



# City Research Online

## City St George's, University of London

**Citation:** Guzmán-Iñigo, J., Sodar, M. A. & Papadakis, G. (2019). Data-based, reduced-order, dynamic estimator for reconstruction of nonlinear flows exhibiting limit-cycle oscillations. *Physical Review Fluids*, 4(11), 114703. doi: 10.1103/physrevfluids.4.114703

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/29234/>

**Link to published version:** <https://doi.org/10.1103/physrevfluids.4.114703>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

# A data-based, reduced-order, dynamic estimator for reconstruction of non-linear flows exhibiting limit-cycle oscillations

Juan Guzmán-Iñigo\*

*Dept. of Mechanical Engineering, Imperial College, London SW7 2AZ, UK*

Markus Sodar and George Papadakis†

*Dept. of Aeronautics, Imperial College, London SW7 2AZ, UK*

## Abstract

We apply a data-based, linear dynamic estimator to reconstruct the velocity field from measurements at a single sensor point in the wake of an aerofoil. In particular, we consider a NACA0012 aerofoil at  $Re = 600$  and  $16^\circ$  angle of attack. Under these conditions, the flow exhibits a vortex shedding limit cycle. A reduced order model (ROM) of the flow field is extracted using proper orthogonal decomposition (POD). Subsequently, a subspace system identification algorithm (N4SID) is applied to extract directly the estimator matrices from the reduced output of the system (the POD coefficients). We explore systematically the effect of the number of states of the estimator, the sensor location, the type of sensor measurements (one or both velocity components), and the number of POD modes to be recovered. When the signal of a single velocity component (in the stream wise or cross stream directions) is measured, the reconstruction of the first two dominant POD modes strongly depends on the sensor location. We explore this behaviour and provide a physical explanation based on the non-linear mode interaction and the spatial distribution of the modes. When however, both components are measured, the performance is very robust, and is almost independent of the sensor location when the optimal number of estimator states is used. Reconstruction of the less energetic modes is more difficult, but still possible. Finally, we assess the robustness of the estimator at off-design conditions, at  $Re = 550$  and  $650$ .

---

\* j.guzman-inigo@imperial.ac.uk

† corresponding author, g.papadakis@imperial.ac.uk

## I. INTRODUCTION

Flow reconstruction (or estimation) from limited measurements has a very wide range of applications, for example in active flow control (drag reduction or enhanced mixing and reaction), cardiovascular flows (extraction of shear stress patterns that are known to affect the behaviour of endothelial cells and the inception of atherosclerosis), optimisation of energy extraction, etc. The area is vast and below only some characteristic approaches/directions are briefly sketched in order to place the present contribution in context.

Due to the large spatial dimensionality of 2D and 3D flows, reduced order models that extract the dominant structures have been naturally employed for flow reconstruction. For example, models based on Proper Orthogonal Decomposition (or POD) [1] have been widely applied. Using a set of sensors embedded in the flow, the objective is to compute the coefficients of POD modes so as to minimise the error (usually defined as the  $\mathcal{L}_2$ -norm) between the true and estimated flow field. Willcox [2] applied an extension of POD, called Gappy POD [3] to find the optimal placement of the sensor points. Yildirim et al (2009) [4] compared Gappy POD as well as a more intuitive approach (placement of sensors at the peaks of modes) and found that the latter works very well. Semaan [5] computed optimal sensor placements using machine learning. Other reduced order modelling strategies, such as the resolvent analysis [6] have been used for estimation of laminar (but also turbulent) flows using limited measurements [7–10]. More recent trends use sparse representation and compressive sensing [11, 12]. The linear stochastic estimation (LSE) techniques of Adrian can also estimate the velocity field from multiple measured observations at nearby points [13–20]. The connection between LSE and POD is detailed in [1].

All the previous approaches are considered as *static* estimators. The coefficients of the reduced order models (or libraries in the case of sparse representation) are obtained by solving a minimisation problem (usually based on least squares, but other norms have also been employed), however the history (dynamic) effect of the coefficients is not considered. For *dynamic* estimation, the Kalman filter is the standard approach for linear problems (refer to the book of Kailath et al [21] for a very lucid and detailed exposition). The filter assumes that the linear system matrices are known, and formulates an optimisation problem that minimises  $\mathcal{L}_2$ -norm between the true and the estimated output. The formulation results into an algebraic Riccati equation that provides the gain matrix of the estimator. This matrix

together with the available measurements are used to force the linear system and extract the estimated states that best reconstruct the output. This approach was used for instance by Gong et al. [22] to successfully estimate the unsteady flow past a cylinder or by Gómez et al. [23] for the flow past an aerofoil. The extended Kalman filter can be used for estimation of non-linear problems [24].

In many practical problems however, the system matrices are unknown, and only input-output data are available. In such cases, a different approach based on system identification algorithms [25] can be applied. These algorithms extract the unknown coefficients of underlying mathematical models using only the available input-output data. System identification has been successfully used to describe the dynamics between one sensor (upstream measurement) to another sensor (downstream measurement) in noise amplifier flows and for linear dynamics. Those models were used to design compensators to effectively estimate and reduce the perturbation field in such flows [26–29]. Guzmán-Iñigo *et al.* [30][31] extended the previous approaches to capture the dynamics between upstream measurements and the *entire* perturbation field in a laminar boundary layer. To do this, the authors applied the “Numerical algorithm for Subspace State-Space System IDentification” (N4SID) [32], an algorithm from the subspace identification family [33]. This group of algorithms extract the matrices of a (linear) state-space estimator model, using only the available input-output data. The algorithm has been applied successfully to many industrial problems (see examples in the book [34]) but its performance on non-linear fluid mechanics problems has yet to be explored (at least to the best of the authors’ knowledge). The connection between N4SID and the Kalman filter is explored in [34].

Dynamic estimation that retains the non-linearity of the underlying process is much more difficult, but progress has been made recently. Loiseau et al [35] use sparse identification to extract an observer that captures the non-linear dynamics of growth and saturation of instabilities in the wake of a circular cylinder. The estimator uses measurements of the lift coefficient and a 2 equation, non-linear, dynamical system is derived. Another approach is based on lifting the original non-linear dynamical system in a higher dimensional space where the evolution is approximately linear (in an uncontrolled system this procedure amounts to numerical approximation of the Koopman operator associated to the non-linear dynamics, see [36]). The discrete empirical interpolation method has also been used to nonlinear model order reduction [37, 38].

While the above non-linear identification approaches are very promising, they are not as well developed as linear ones. The central objective of the present paper is to assess the performance of linear dynamical identification tools for the estimation of non-linear flows. We consider the estimation of the fully developed vortex shedding state, i.e. the limit-cycle, using a single sensor placed at different locations in the wake of a NACA 0012 aerofoil.

The paper is organised as follows. In section II we provide the theoretical framework of the identification, section III describes the flow configuration, while results are presented in sections IV and V. A performance analysis in off-design conditions is provided in section VI. We conclude in section VII.

## II. THEORETICAL FRAMEWORK

In this section we present a linear model to estimate the dynamics of a fully non-linear flow. The equations governing a reduced-order model based on the perturbation dynamics around the time-average (mean) flow are introduced and subsequently used to justify the structure of the data-based, linear dynamic estimator. Finally, a linear system identification algorithm is introduced that can determine efficiently the matrices of the estimator directly from the observed input-output data via a statistical learning process.

### A. Reduced Order Model based on the perturbation dynamics about a mean flow

We consider the incompressible Navier-Stokes equations

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p - \frac{1}{Re} \nabla^2 \mathbf{u} = 0, \quad \nabla \cdot \mathbf{u} = 0, \quad (1)$$

where  $\mathbf{u}$  is the velocity vector,  $p$  the pressure and  $Re$  the Reynolds number. The flow variables are decomposed into a time-average (denoted by overbar) and a fluctuating part (denoted by prime); e.g.  $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$ . Applying the time-averaging operation to Eq. (1) results in

$$\bar{\mathbf{u}} \cdot \nabla \bar{\mathbf{u}} + \nabla \bar{p} - \frac{1}{Re} \nabla^2 \bar{\mathbf{u}} = -\overline{\mathbf{u}' \cdot \nabla \mathbf{u}'}, \quad \nabla \cdot \bar{\mathbf{u}} = 0, \quad (2)$$

which governs the steady, mean flow. The equations governing the fluctuation part are obtained by subtracting Eq. (2) from Eq. (1) to give

$$\frac{\partial \mathbf{u}'}{\partial t} + \bar{\mathbf{u}} \cdot \nabla \mathbf{u}' + \mathbf{u}' \cdot \nabla \bar{\mathbf{u}} + \nabla p' - \frac{1}{Re} \nabla^2 \mathbf{u}' = \overline{\mathbf{u}' \cdot \nabla \mathbf{u}'} - \mathbf{u}' \cdot \nabla \mathbf{u}', \quad (3a)$$

$$\nabla \cdot \mathbf{u}' = 0, \quad (3b)$$

which has been arranged so that the left-hand side is linear in the fluctuation variables. Treating the nonlinearities as a forcing term, Eq. (3) can be rewritten in the classical state-space form

$$\mathcal{Q} \frac{\partial \mathcal{X}}{\partial t} + \mathcal{A} \mathcal{X} = \mathcal{F}, \quad (4)$$

where

$$\mathcal{X} = \begin{pmatrix} \mathbf{u}' \\ p' \end{pmatrix}, \quad (5a)$$

$$\mathcal{Q} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & 0 \end{pmatrix}, \quad (5b)$$

$$\mathcal{A} = \begin{pmatrix} \bar{\mathbf{u}} \cdot \nabla (\cdot) + (\cdot) \cdot \nabla \bar{\mathbf{u}} - Re^{-1} \nabla^2 \nabla (\cdot) & \\ \nabla \cdot (\cdot) & 0 \end{pmatrix}, \quad (5c)$$

$$(5d)$$

and

$$\mathcal{F} = \begin{pmatrix} \overline{\mathbf{u}' \cdot \nabla \mathbf{u}'} - \mathbf{u}' \cdot \nabla \mathbf{u}' \\ 0 \end{pmatrix}. \quad (6)$$

We now introduce a divergence-free orthonormal basis  $\{\Phi_i\}_{1\dots m}$ , satisfying

$$\nabla \cdot \Phi_i = 0 \quad \text{and} \quad \langle \Phi_i(\mathbf{x}), \Phi_j(\mathbf{x}) \rangle = \delta_{ij}, \quad i = 1, 2, \dots, m \quad (7)$$

where  $\langle \cdot, \cdot \rangle$  is an appropriate scalar product (to be defined later) and  $\delta_{ij}$  is the Kronecker delta. The velocity field can then be projected onto this basis according to

$$\mathbf{u}'(\mathbf{x}, t) = \sum_{i=1}^m y_i(t) \Phi_i(\mathbf{x}) \quad i = 1, 2, \dots, m \quad (8a)$$

$$y_i(t) = \langle \Phi_i(\mathbf{x}), \mathbf{u}'(\mathbf{x}, t) \rangle. \quad (8b)$$

We define a state vector given by the first  $k$  coefficients of the previous expansion  $\mathbf{Y} = [y_1, y_2, \dots, y_k]^\top$  (where  $^\top$  denotes the transpose) with a corresponding reduced basis  $\mathbf{U} =$

$[\Phi_1, \Phi_2, \dots, \Phi_k]$ . Finally, we perform a Galerkin projection of Eq. (4) onto the subspace spanned by  $\mathbf{U}$  to obtain

$$\frac{d\mathbf{Y}}{dt} + \mathbf{A}'\mathbf{Y}(t) = \mathbf{F}'(t) + \boldsymbol{\epsilon}(t), \quad (9a)$$

where

$$\mathbf{A}'_{ij} = \langle \Phi_i, \mathcal{A}\Phi_j \rangle, \quad \mathbf{F}'_i = \langle \Phi_i, \mathcal{F} \rangle \quad \text{and} \quad \epsilon_i = \sum_{j=k+1}^m y_j \langle \Phi_i, \mathcal{A}\Phi_j \rangle. \quad (9b)$$

Note that  $\boldsymbol{\epsilon}$  represents the error due to the truncation of the expansion ( $k < m$ ).

## B. Equations of the linear dynamic estimator

We now consider a flow estimator of the form

$$\frac{d\mathbf{X}_e}{dt} + \mathbf{A}'_s\mathbf{X}_e(t) = \mathbf{L}'\mathbf{s}(t), \quad (10a)$$

$$\mathbf{Y}_e(t) = \mathbf{C}'\mathbf{X}_e(t), \quad (10b)$$

$$\mathbf{s}(t) = \mathbf{C}'_s\mathbf{Y}(t) + \mathbf{g}(t), \quad (10c)$$

where  $\mathbf{X}_e \in \mathbb{R}^{N_x}$  is the estimator state vector,  $\mathbf{Y}_e \in \mathbb{R}^k$  is the estimator output and  $\mathbf{s} \in \mathbb{R}^p$  contains measurements from  $p$  sensor points placed in the flow,  $\mathbf{s} = [s_1, s_2, \dots, s_p]^\top$ . In the present study, we employ sensors that are linear with respect to the velocity field, i.e.  $s_j(t) = \mathcal{C}_j\mathbf{u}(t) + g_j(t)$ , where  $g_j(t)$  is the noise that corrupts the measurements. The linear operator  $\mathcal{C}_j$  is projected onto the reduced-order basis as  $\mathbf{C}_{s,ji} = \langle \Phi_i, \mathcal{C}_j \rangle$ .

We seek to construct the estimator so that the approximate output  $\mathbf{Y}_e(t)$  is as close as possible to the true output  $\mathbf{Y}(t)$ . Intuitively, this will be achieved if the measurements from the sensor  $\mathbf{L}'\mathbf{s}(t)$  correctly represent the non-linear term  $\mathbf{F}'(t)$ . Note that the sensor may also account for the truncation error  $\boldsymbol{\epsilon}$ , if this is non-negligible.

The equation governing the estimation error, defined as  $\mathbf{Z} = \mathbf{Y} - \mathbf{Y}_e$ , is obtained by subtracting Eq. (10) from Eq. (9) to yield

$$\frac{d\mathbf{Z}}{dt} + \mathbf{C}'\mathbf{A}'_s\mathbf{C}'^\dagger\mathbf{Z}(t) = \mathbf{F}^{\text{err}}(t), \quad (11a)$$

with

$$\mathbf{F}^{\text{err}}(t) = \mathbf{A}^{\text{err}}\mathbf{Y}(t) + \mathbf{F}'(t) - \mathbf{C}'\mathbf{L}'\mathbf{g}(t) + \boldsymbol{\epsilon}(t), \quad \text{and} \quad \mathbf{A}^{\text{err}} = (\mathbf{C}'\mathbf{A}'_s\mathbf{C}'^\dagger - \mathbf{A}' - \mathbf{C}'\mathbf{L}'\mathbf{C}'_s). \quad (11b)$$

The relation  $\mathbf{X}_e = \mathbf{C}'^\dagger \mathbf{Y}_e$  was used to obtain Eq. (11) where  $\mathbf{C}'^\dagger$  denotes the Moore-Penrose generalised matrix inverse [39]. Eq. (11) shows that the term  $\mathbf{F}^{\text{err}}$  must be small for the error to be small and that the eigenvalues of  $\mathbf{A}'_s$  and  $\mathbf{A}^{\text{err}}$  must be stable.

With the remaining article pertaining to system identification methods, it is more convenient to express the estimator in a discrete-time framework. In the discrete-time domain, the mapping of the state-vector  $\mathbf{Y}_e$  from time  $t$  (index  $n$ ) to  $t + \Delta t$  (index  $n + 1$ ) reads

$$\mathbf{X}_e(n + 1) = \mathbf{A}_s \mathbf{X}_e(n) + \mathbf{L} \mathbf{s}(n), \quad (12a)$$

$$\mathbf{Y}_e(n) = \mathbf{C} \mathbf{X}_e(n), \quad (12b)$$

$$\mathbf{s}(n) = \mathbf{C}_s \mathbf{Y}(n) + \mathbf{g}(n), \quad (12c)$$

where matrix  $\mathbf{L} = \int_0^{\Delta t} \exp[-\mathbf{A}'_s(\Delta t - \tau)] \mathbf{L}' d\tau$  is associated with the discrete driving term,  $\mathbf{A}_s = \exp(-\mathbf{A}'_s \Delta t)$  denotes the evolution matrix over a time interval  $\Delta t$ ,  $\mathbf{C} = \mathbf{C}'$  and  $\mathbf{C}_s = \mathbf{C}'_s$  (see also [40]).

### C. System identification based on subspace techniques

We seek to obtain a linear, time-invariant (LTI) multiple-input-multiple-output (MIMO) system, such as the one given in Eq. (12), from a sequence of observed input-output data. Subspace identification algorithms can accomplish this task. More specifically, such algorithms consider the state-space representation of a stochastic linear system written in the following *process form*

$$x(n + 1) = \mathcal{A}x(n) + \mathcal{B}u(n) + w(n) \quad (13a)$$

$$y(n) = \mathcal{C}x(n) + \mathcal{D}u(n) + v(n) \quad (13b)$$

where  $y(n) \in \mathbb{R}^{N_y}$ ,  $x(n) \in \mathbb{R}^{N_x}$ ,  $u(n) \in \mathbb{R}^{N_u}$ ,  $w(n) \in \mathbb{R}^{N_x}$ ,  $v(n) \in \mathbb{R}^{N_y}$  are the system output, state, input, state noise, and output measurement noise, respectively. The matrices  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$  and  $\mathcal{D}$  have appropriate dimensions. The noise covariances are defined as

$$\mathcal{E} \left\{ \begin{pmatrix} w(j) \\ v(j) \end{pmatrix} \begin{pmatrix} w(i) \\ v(i) \end{pmatrix}^\top \right\} = \begin{pmatrix} \mathcal{Q} & \mathcal{S} \\ \mathcal{S}^\top & \mathcal{R} \end{pmatrix} \delta_{ij} \quad (14)$$

where  $\mathcal{E}\{x\}$  stands for the expectation operator.

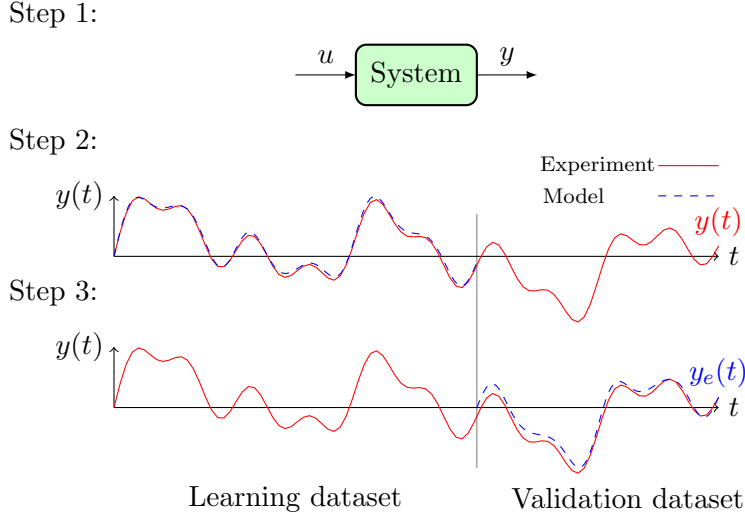


FIG. 1: Procedural steps of system identification techniques. Step 1: Numerical simulations/experiments are run and input ( $u$ ) and output data ( $y$ ) are acquired. Step 2: The model coefficients are computed by maximising the fit between the output of the system and the prediction of the model for part of the available data (training or learning data set). Step 3: The performance of the model is assessed on a different dataset (testing or validation dataset).

Subspace identification methods extract the system matrices  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$  and  $\mathcal{D}$ , as well as the covariance matrices  $\mathcal{Q}$ ,  $\mathcal{S}$ , and  $\mathcal{R}$  from a set of input-output measurements. A comprehensive description of these methods is given in the review paper of Qin [33] and the book of Van Overschee and De Moor [34]. In this study, the N4SID algorithm [32, 41] implemented in MATLAB has been used to obtain all the matrices. Eq. (13) intends to mimic Eq. (12) and, thus, the matrix  $\mathcal{D}$  and the noise components  $w(n)$  and  $v(n)$  are set to zero. The input  $u$  corresponds to sensor measurements  $\mathbf{s}$  and the output  $y$  to the coefficients of the selected modes  $\Phi_i$ , which are taken to be the POD modes.

The N4SID algorithm works in two steps. In the first step, the state sequence  $x(n)$  as well as the order of the system  $N_x$ , are extracted from the input-output data. During this step the (extended) observability matrix, with rank equal to  $N_x$ , is also obtained. The crucial component in this step is the singular value decomposition of a weighted matrix. In the second step, the matrices  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{C}$ , are calculated (to within a similarity transformation) from the system states  $x(n)$  by least squares. We note that the order  $N_x$  can be specified by the user, or extracted directly from the data (we explore both options later in Sec. V).

There are other subspace identification methods, which are related with N4SID. This relation is established by the unifying identification theorem [34, 41].

The application of the aforementioned system identification technique to our problem requires three procedural steps which are depicted in Fig. 1. In the first step, numerical simulations of the full problem are carried out and the inputs (velocity measurements at selected points in the wake) and outputs (coefficients of POD modes) of the estimator are extracted. In the second step, a sub-sample of the extracted data set, referred to as the learning dataset, is processed to determine the system matrices of the identified model. In a third step, a different part of the data set, known as the testing dataset, is used to drive the identified system, and the output  $y_e(t)$  produced by the estimator is compared to the measured true output  $y(t)$ .

### III. FLOW CONFIGURATION AND NUMERICAL APPROACH

The general formalism outlined in the previous section is applied to a flow configuration that is now detailed. We consider the two-dimensional flow around a NACA0012 aerofoil with chord  $c$ , that forms an angle of attack  $\alpha$  with the uniform approaching velocity  $U_\infty$ . We choose  $c$  and  $U_\infty$  as the reference scales for length and velocity respectively. The Reynolds number, defined as  $Re = U_\infty c / \nu$ , is set to  $Re = 600$  and  $\alpha = 16^\circ$ . These conditions correspond to an unsteady laminar flow that reaches a limit cycle (the critical Reynolds number at this  $\alpha$  is  $Re_c \approx 400$ , refer to [42]).

Fig. 2 (left) illustrates the whole computational domain. The Cartesian coordinate system  $(x, y)$  is centred at the point where the leading edge of the aerofoil with zero angle of attack would have been placed. The domain size is 40 units in the  $x$ -direction, with the wake region extending 25 units from the trailing edge. In the  $y$ -direction, the domain size is 15 units above and below the trailing edge.

The numerical simulations were performed with Star-CCM+ Version 13.02. The governing equations are discretised using the finite-volume method. The domain is spatially discretised with an unstructured polyhedral mesh, while an additional prism layer is inserted around the aerofoil to better resolve the near wall region. The mesh is composed of 71,209 cells in total. The discrete system of equations is integrated in time using a second-order implicit scheme. The normalised tolerance for all equations is set to  $10^{-4}$ .

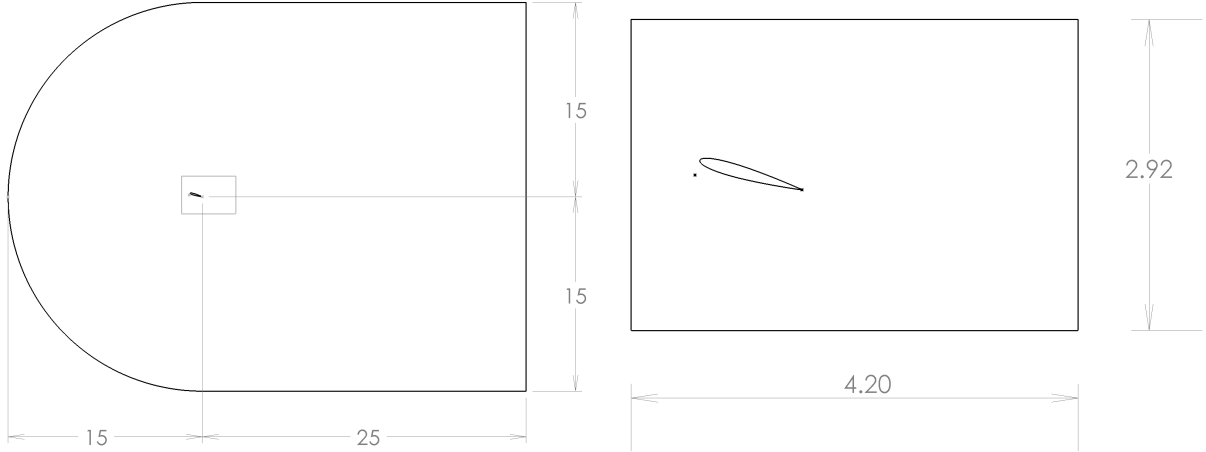


FIG. 2: (Left) Computational domain and (right) control window for data extraction.

The velocity is set to  $\mathbf{u} = (1, 0)$  at the inlet, no-slip condition is applied on the aerofoil wall, and a pressure boundary condition is imposed at the outlet.

The simulation starts from a zero field, and after a transient period the flow develops into a vortex shedding limit-cycle, with fundamental frequency  $f = 0.61$  (the value matches perfectly with the result of [42]). The time-step is  $\Delta t_{DNS} = 0.01$ , which corresponds to more than 160 steps per period. We have also compared the spectra at different points in the wake with those of [42] and again good matching was found (results not presented for brevity). The simulation continues for a total time of  $t = 80$ , which corresponds to 50 vortex shedding periods.

An instantaneous snapshot of the vorticity and velocity fields is depicted in Fig. 3. Due to the high angle of attack, the flow separates at the leading edge in the suction side, and vorticity is periodically shed in the wake. In the pressure side, the flow remains attached; this side also periodically releases vortices in the wake. The vortex pattern can be clearly observed in both figures. Due to the flow asymmetry, the vortices emitted from the two sides have different strengths.

In order to compute the time-averaged flow, an initial transient period (of length  $T = 24$ ) is discarded, and a dataset of length  $T = 56$  (corresponding to 43 vortex shedding periods) is used. The resulting mean flow is plotted in Fig. 4. A low-speed recirculation region, with approximate length 1.5, forms in the wake. This region is composed of two bubbles with opposite circulations: (i) a large one encompassing most of the suction side and (ii) a much smaller one close to the trailing edge.

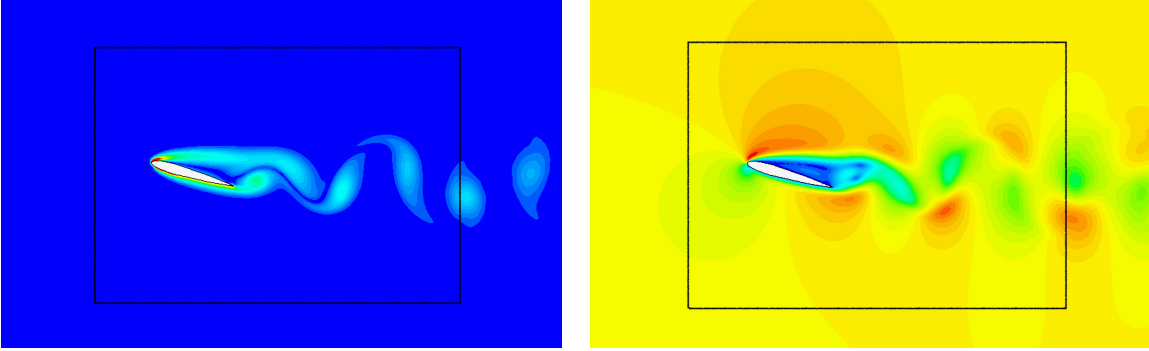


FIG. 3: Snapshots of the (left) vorticity field and (right) velocity magnitude. The solid black line denotes the outline of the control window.

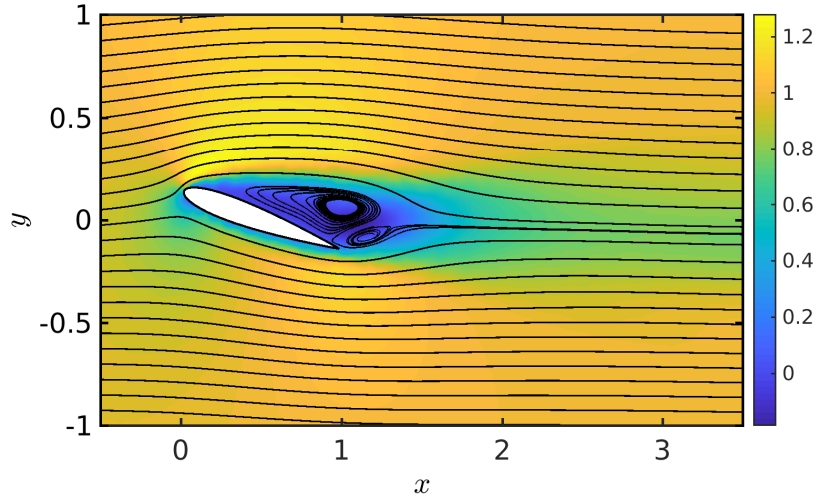


FIG. 4: Streamwise component of the mean flow velocity. Black lines denote streamlines.

In order to compute the estimator outlined in Sec. II, velocity data are extracted from points inside the control window shown in Fig. 2(right). The window is sampled with 100 equidistant points with spacing  $\Delta x = 0.042$  and  $\Delta y = 0.0292$  in the  $x$  and  $y$ -directions, respectively. The points located inside the aerofoil are discarded by the software (this applies to 63 points and consequently 9937 nodes output velocity data). The mean flow is subtracted and only the fluctuations around the mean, i.e. the oscillatory part, are processed.

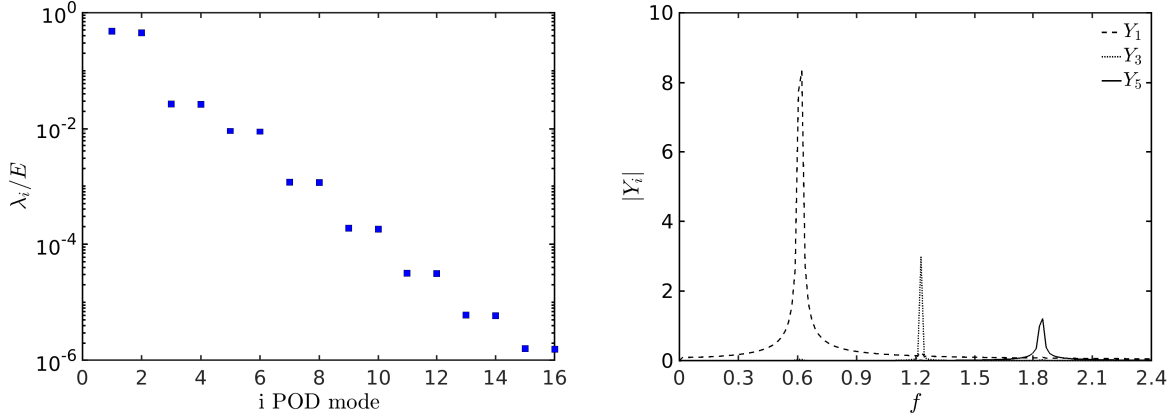


FIG. 5: (Left) First 16 POD eigenvalues  $\lambda_i$  of the correlation matrix and (right) spectrum of the first, third and fifth POD coefficients.

## IV. RESULTS

### A. Reduced-order data

The large spatial dimensionality of the data is beyond the current capabilities of most system identification algorithms and needs to be reduced. Let us consider a sequence of  $m$  velocity snapshots  $\{\mathbf{V}_{\text{snap}(n)}\}_{n=1\dots m}$  extracted from the simulation. The proper orthogonal decomposition (POD) enables us to compute a ranked orthonormal basis,  $\{\Phi_i\}_{i=1\dots m}$ , that satisfy the conditions given by Eqs. (7). Any velocity field  $\mathbf{V}$  can then be projected onto this first  $k$  modes to produce the approximate flow field  $\mathbf{V}'$  (refer to Eq. (8)).

The POD decomposition guarantees that the norm-2 of the error  $\|\mathbf{V} - \mathbf{V}'\|^2 = \langle \mathbf{V} - \mathbf{V}', \mathbf{V} - \mathbf{V}' \rangle$  is minimal for the set of  $m$  snapshots and a given order  $k$ . The scalar product between two vectors  $\langle \mathbf{u}^1, \mathbf{u}^2 \rangle$ , where  $\mathbf{u}^1 = (u^1, v^1)$  and  $\mathbf{u}^2 = (u^2, v^2)$ , is defined as

$$\langle \mathbf{u}^1, \mathbf{u}^2 \rangle = \int \int_{\Omega} (u^1 u^2 + v^1 v^2) d\Omega. \quad (15)$$

When  $\mathbf{u}^1 = \mathbf{u}^2$ , the norm is equal to (twice) the kinetic energy in the domain. Based on this scalar product, the POD basis captures the maximum proportion of kinetic energy in the domain  $\Omega$  for any given order  $k$ .

The POD decomposition was applied to the oscillatory part of the same snapshots used to compute the mean flow. Fig. 5(left) shows the eigenvalues of the correlation matrix (normalised with the total energy of the fluctuating field) for the first 16 POD modes.

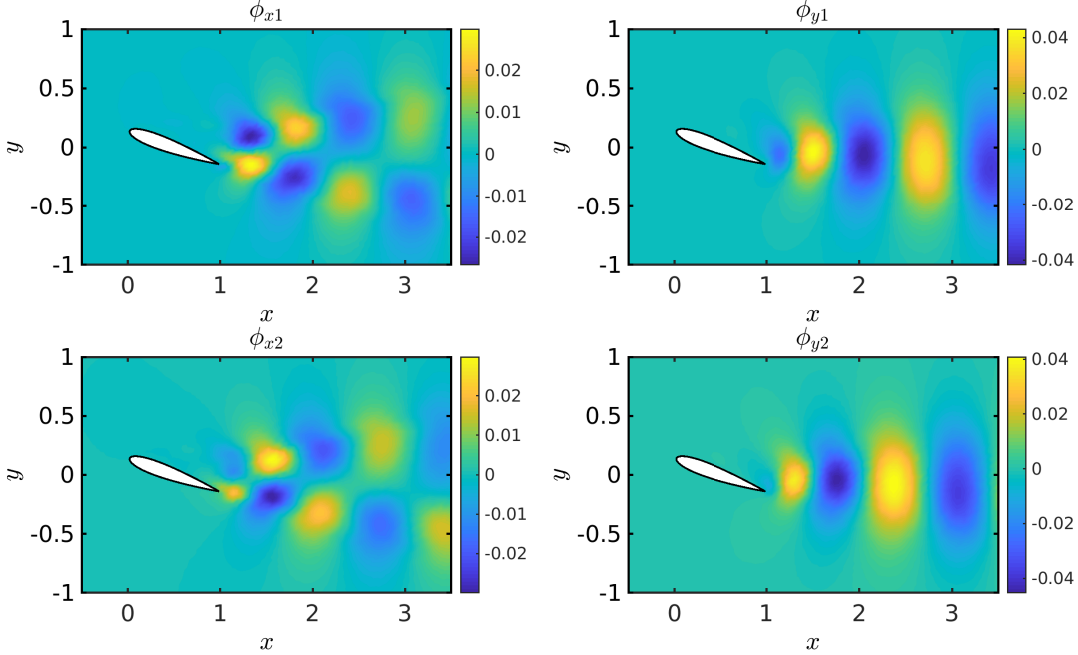


FIG. 6: Contours of the (left column) streamwise and (right column) cross-stream velocity components of the (top) first and (bottom) second POD-modes.

These eigenvalues represent the energy content carried by each mode. Over 92% of the oscillatory kinetic energy is captured by the first 2 modes and over 97% by the first 4. Due to the convective nature of vortex shedding motion, POD modes arise in pairs.

The spectra of the POD coefficients, shown in Fig. 5(right), indicate that the first (and second) POD modes are associated with the fundamental vortex-shedding frequency ( $f = 0.61$ ), while the third (and fourth) and fifth (and sixth) modes with the first and second harmonics, respectively. These harmonics are produced by non-linear interactions in the wake. For example, the first harmonic is produced by the interaction of the first two modes (more details can be found in [43]). The placement of the sensor point in relation to the spatial location of these interactions affects the performance of the estimator, as will be shown later.

Fig. 6 depicts the streamwise and cross-stream velocity components of the first and second mode. This pair clearly captures the large coherent structures corresponding to vortex shedding. Note that the two modes have very similar structure, but are shifted in space. The coefficients are also shifted in time by 90 deg and this combination results in the propagation of structures in the wake. Fig. 7 depicts the velocity components of the third and fourth

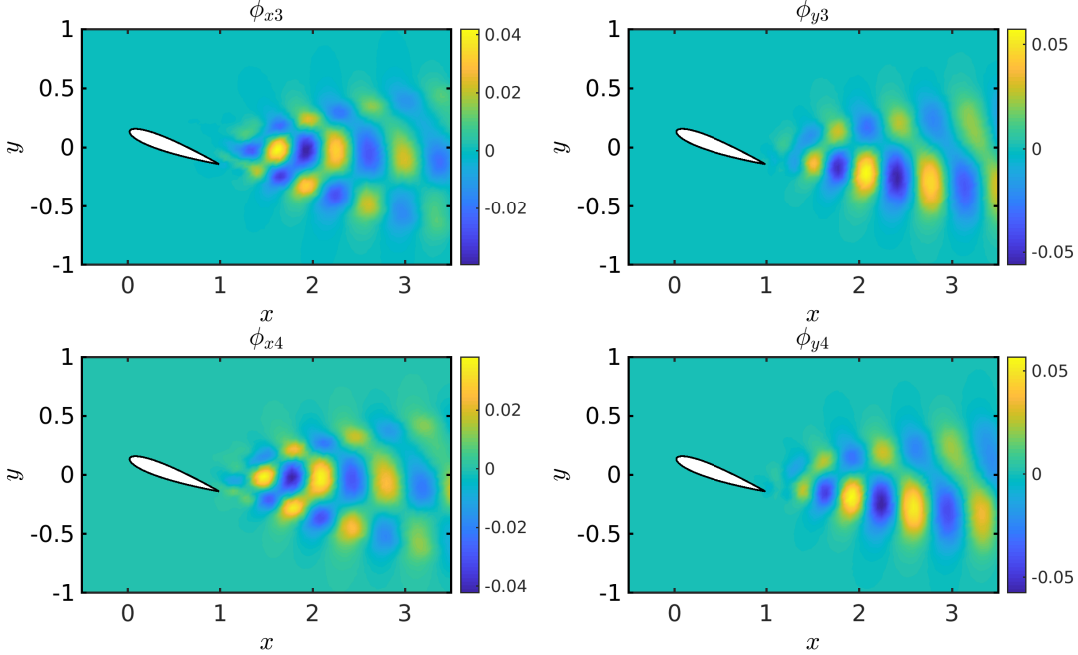


FIG. 7: Contours of the (left column) streamwise and (right column) cross-stream velocity components of the (top) third and (bottom) fourth POD-modes.

POD modes. These structures are smaller compared to those of the first and second mode and are also shifted in space.

For the application of the system identification algorithm, the time-evolving POD coefficients corresponding to the  $k$  most energetic modes are arranged in the reduced order state vector  $\mathbf{Y} = [y_1, y_2, \dots, y_k]^\top$  and constitute the output of our system.

## B. System identification and validation

In this section, we seek to build a dynamical system (with the structure given by Eq. (12)) that estimates the evolution of  $\mathbf{Y}(n)$  from local measurements. We consider a sensor measuring the streamwise and cross-stream components of the oscillatory velocity  $s(n) = [u(n), v(n)]^\top$  at the point  $(x_s, y_s) = (1.82, -0.04)$ . Note that for this first application of the identification algorithm, the location of the measuring point is arbitrary. A detailed study of the effect of the sensor position on the quality of the identification is provided later in Sec. V.

A time-segment of the simulation, known as learning dataset, is used together with the

identification algorithm N4SID [32] to estimate the unknown matrices of system (12). The reduced-order model is determined with  $k = 4$  POD modes and represents the evolution of the vortex-shedding mode and its first harmonic. The number of states of the estimator is set to  $N_x = k = 4$ . The length of the learning dataset is  $m_{LD} = 800$  snapshots, with time step  $\Delta t = 0.01$ , yielding a physical time  $T_{LD} = 8$ , which corresponds to 5 vortex-shedding periods.

Fig. 8 shows a comparison of the true and estimated POD coefficients for the learning dataset. The top plot shows the measurements from the input sensor and the time-evolution of the four POD coefficients is illustrated underneath. For the selected parameters, the algorithm accurately estimates the evolution of the coefficients over the entire learning dataset.

To quantify the performance of the estimator, we use the fit between the true  $i$ -th POD coefficient extracted from the DNS and the one predicted defined as

$$\text{FIT}_i[\%] = 100 \times \left( 1 - \frac{\|y_i(t) - y_{i,e}(t)\|}{\|y_i(t) - \overline{y_i(t)}\|} \right), \quad (16)$$

where  $\|\cdot\|$  denotes the 2-norm of a vector. The vector of fits obtained for the learning dataset for the first four POD coefficients is  $\text{FIT}[\%] = [98.77, 97.67, 97.27, 98.07]$ .

For the learning dataset, the good agreement between the true POD coefficients and those predicted by the model is expected since the model is obtained by minimising the error  $(\mathbf{Y} - \mathbf{Y}_e)$  over that period of time. The validity of the identified model must be confirmed using a different time-segment of the simulation. Fig. 9 illustrates the performance of the model on a validation dataset. The model states  $\mathbf{X}_e$  are initialised using two different values: (i) the actual reduced-order state of the system  $\mathbf{X}_e = \mathbf{C}^\dagger \mathbf{Y}$  and (ii)  $\mathbf{X}_e = \mathbf{0}$ . When initialised by the last known state of the system, the prediction of the algorithm accurately matches the evolution of the true POD coefficients over the whole dataset. On the other hand, a transient period of length about  $T = 4$  ( $\approx 2.4$  periods) is required before the model predictions match true results when the model is initialised with  $\mathbf{X}_e = \mathbf{0}$ . After the short transient, the coefficients are correctly predicted. The vector of fits obtained for the validation dataset is  $\text{FIT}[\%] = [98.81, 97.61, 97.18, 98.04]^\top$  when initialised with the actual state.

Eq. (8) allows us to reconstruct the full velocity field from the identified reduced-order model. Fig. 10 depicts the original and reconstructed streamwise component of the velocity field at two arbitrary time instants for the testing dataset. The vortex shedding is clearly

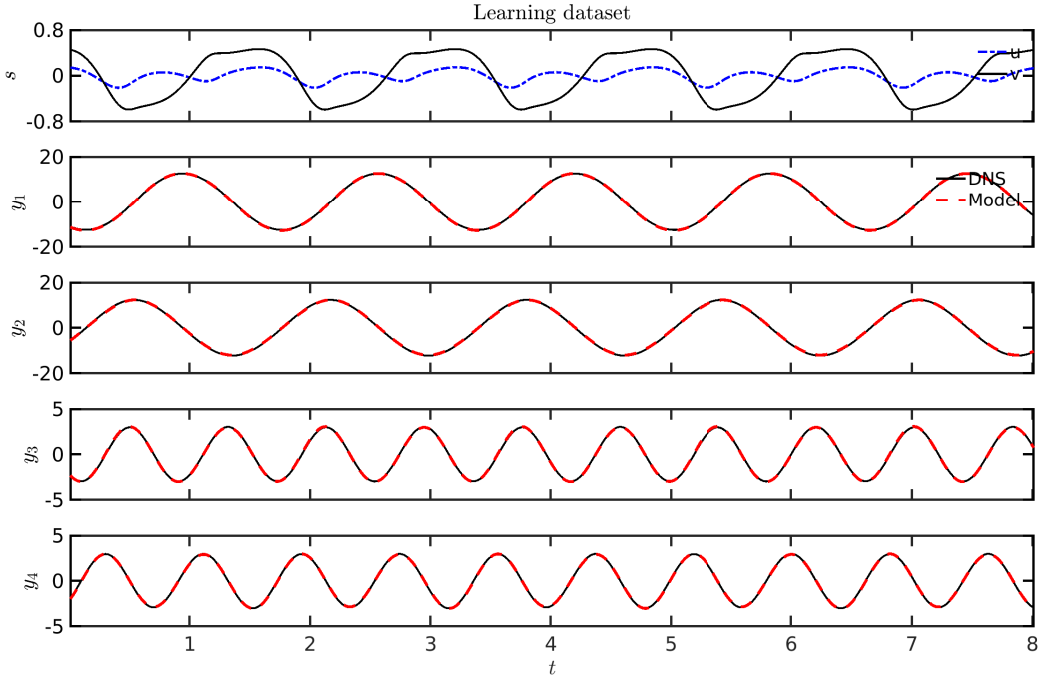


FIG. 8: Learning dataset: (top) the measurement signal  $s(t)$  and (bottom) first four POD coefficients  $y_i(t)$  obtained by projecting the flow field onto the POD modes (solid black) and predicted by the model (dashed red).

captured by the model and the reconstructions at both instants appear almost identical.

We can evaluate the fit between the true perturbation velocity field at each point (computed from the DNS) and the value predicted by the model in the same way as we did for the POD coefficients in Eq. (16). We simply need to replace  $y_i(t)$  and  $y_{i,e}(t)$  by the true and estimated velocity values respectively at a given point. Fig. 11 shows contour plots of the FIT[%] for the validation dataset. A good agreement is found in the whole domain, with FIT values ranging from 75% to 100%. The areas with the largest mismatch are in the wake of the aerofoil and coincide with the regions of largest amplitude of the second harmonic, which is not considered in the current estimator model.

It is instructive to study also the Bode plots of the transfer function between the input  $s = [u, v]^T$  and the output  $\mathbf{Y}_e(t) = [y_{1,e}(t), y_{2,e}(t), y_{3,e}(t), y_{4,e}(t)]^T$ . The transfer functions are obtained by taking the Fourier transform of Eq. (10). The result is  $\hat{\mathbf{Y}}_e(\omega) = \mathbf{C}'(\omega\mathbf{I} + \mathbf{A}'_s)^{-1} \mathbf{L}'\hat{\mathbf{s}}(\omega)$  or in expanded form

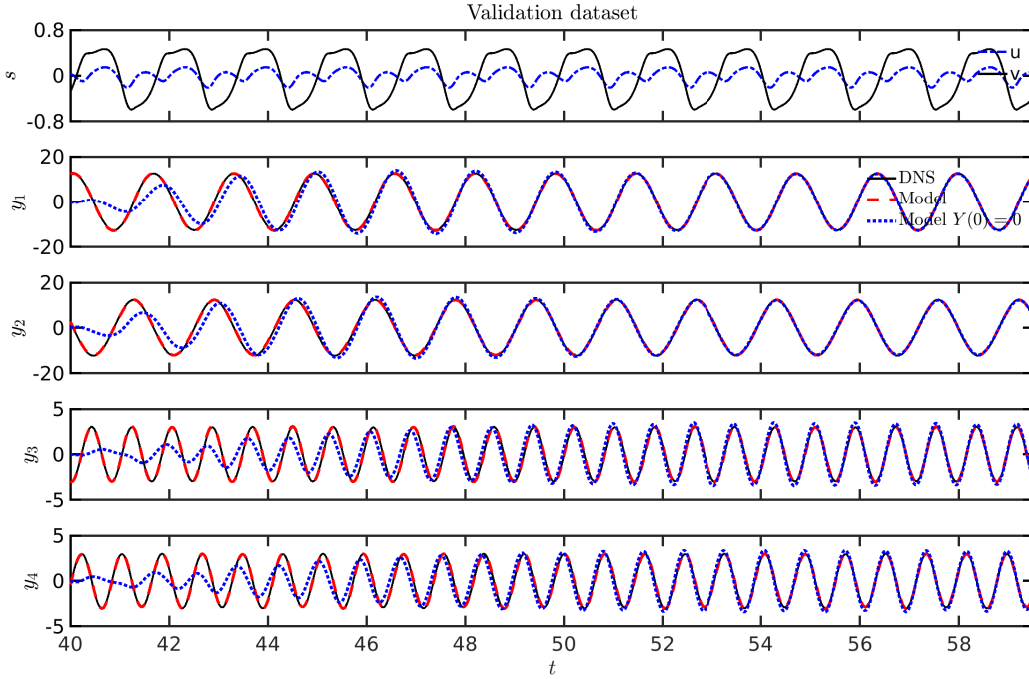


FIG. 9: Validation dataset: Performance of the system-identified model. (Top) Input data  $s$  and (bottom) comparison between the DNS (black solid) and model prediction for four POD coefficients  $y_i$ . The model is initialised by  $\mathbf{X}_e = \mathbf{C}^\dagger \mathbf{Y}$  (red dashed) and  $\mathbf{X}_e = \mathbf{0}$  (blue dotted) at  $t = 40$ .

$$\begin{pmatrix} \hat{y}_{1,e}(\omega) \\ \hat{y}_{2,e}(\omega) \\ \hat{y}_{3,e}(\omega) \\ \hat{y}_{4,e}(\omega) \end{pmatrix} = \begin{pmatrix} M_{u \rightarrow y_1}(\omega) & M_{v \rightarrow y_1}(\omega) \\ M_{u \rightarrow y_2}(\omega) & M_{v \rightarrow y_2}(\omega) \\ M_{u \rightarrow y_3}(\omega) & M_{v \rightarrow y_3}(\omega) \\ M_{u \rightarrow y_4}(\omega) & M_{v \rightarrow y_4}(\omega) \end{pmatrix} \begin{pmatrix} \hat{u}(\omega) \\ \hat{v}(\omega) \end{pmatrix} \quad (17)$$

where the hat ( $\hat{\cdot}$ ) denotes the Fourier-transformed variable. Fig. 12 shows the Bode plots (gain and phase) of  $M_{u \rightarrow y_1}(\omega)$  and  $M_{u \rightarrow y_2}(\omega)$  (left column) as well as  $M_{v \rightarrow y_1}(\omega)$  and  $M_{v \rightarrow y_2}(\omega)$  (right column). The estimator has the form of a second order filter, that has high gain close to the frequency of the 1st and 2nd mode, and damps all the other frequencies. At the mode frequency 0.61, the gain from  $v$  is almost twice as large compared to  $u$ . This result together with the spectra of the input signal (plotted in Fig. 13) indicate that it is mainly the  $v$  component that provides the information to extract  $y_{1,e}$  and  $y_{2,e}$  at the selected point. Interestingly, the peak value is not exactly at the frequency 0.61, but slightly displaced.

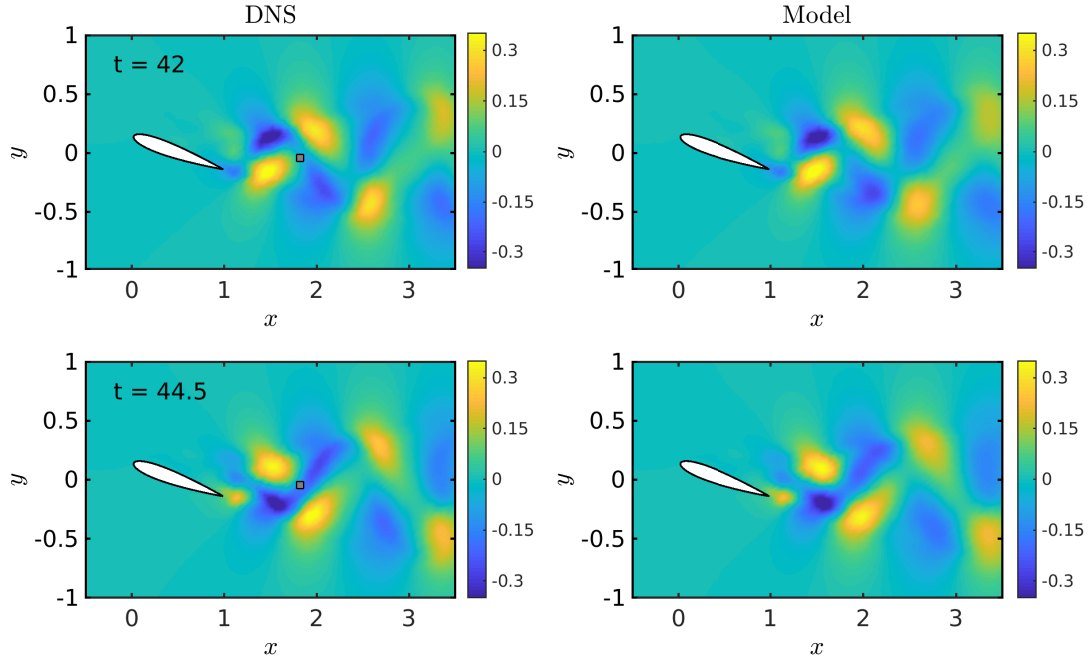


FIG. 10: Streamwise component of the true oscillatory velocity (left) obtained from the DNS and (right) model prediction. The grey squares represent the position of the estimation sensor. See supplementary movie 1.

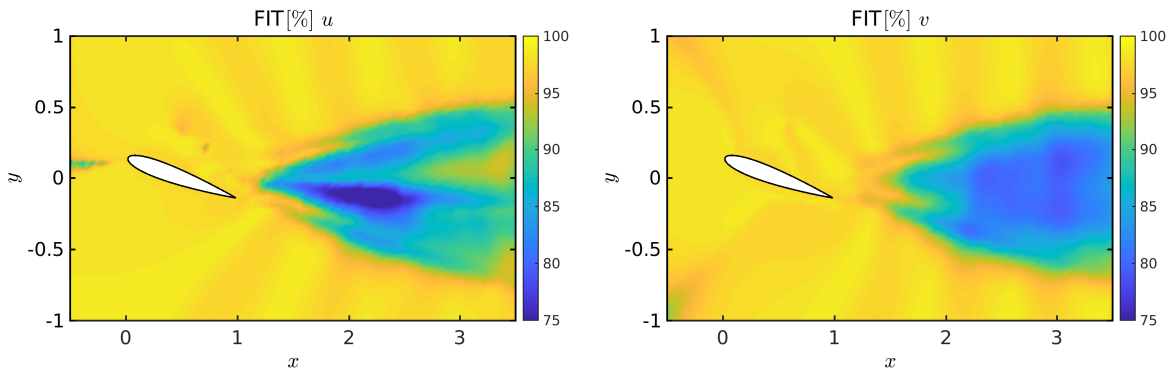


FIG. 11: Goodness of fit FIT[%] between the actual oscillatory velocity field and the field estimated by the model at each point: (left) streamwise and (right) cross-stream components of the velocity.

From the phase plot (bottom row) we note also that the estimator captures the correct 90 deg phase difference between  $y_{1,e}(t)$  and  $y_{2,e}(t)$  at the frequency of vortex shedding.

Fig. 14 shows the Bode plots for the third and fourth modes  $M_{v \rightarrow y_3/y_4}(\omega)$  (left column) and  $M_{v \rightarrow y_3/y_4}(\omega)$  (right column). Again the shape is similar to a second order filter, centred

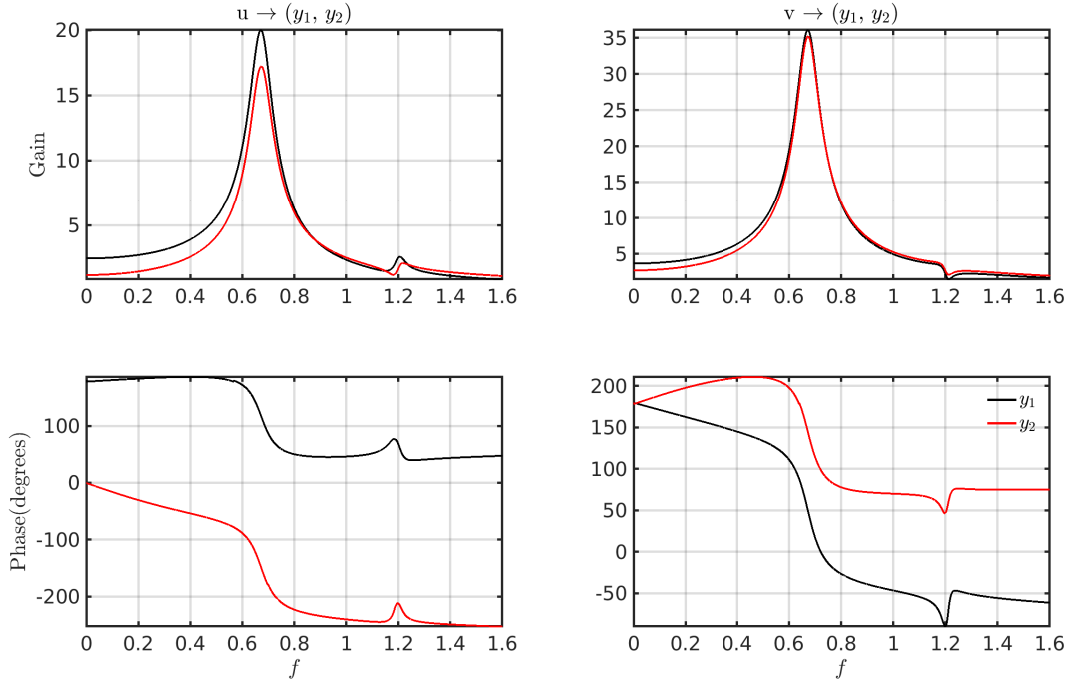


FIG. 12: Bode plots (gain and phase) of the transfer functions from the two inputs ( $u, v$ ) to the two outputs (first and second POD coefficients).

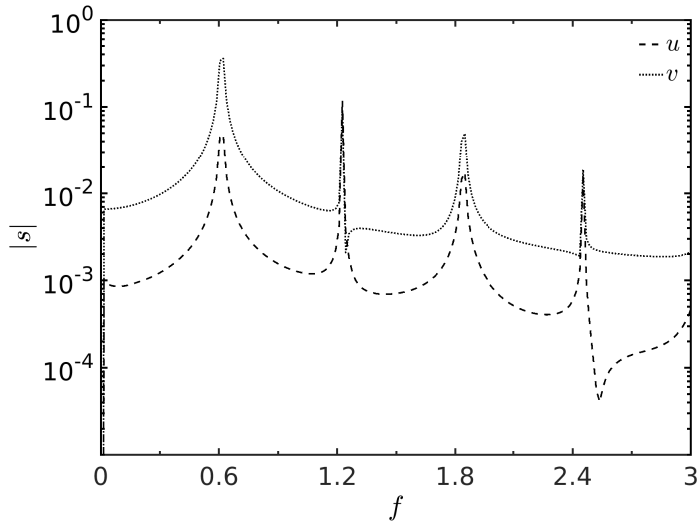


FIG. 13: Spectra of the streamwise  $u$  and cross-stream  $v$  components of the oscillatory velocity composing the input  $s = [u, v]^T$ .

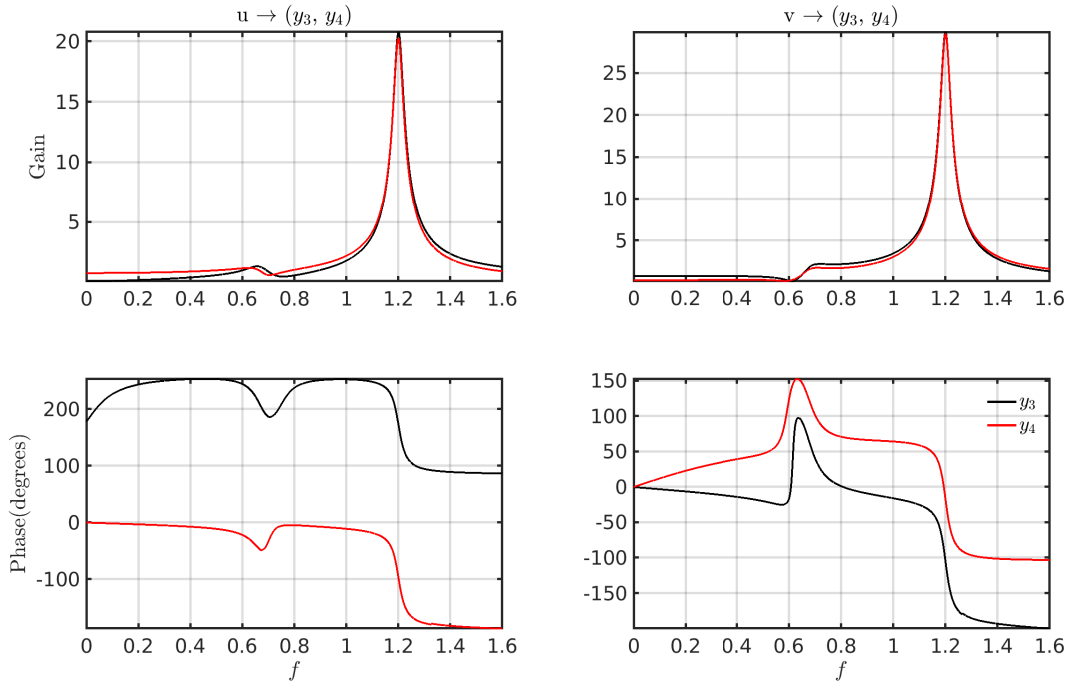


FIG. 14: Bode plots (gain and phase) of the transfer functions from the two inputs ( $u, v$ ) to the two outputs (third and fourth POD coefficients).

at the first harmonic, and all the other frequencies are suppressed. Note that now both  $u$  and  $v$  contribute equally to the extraction of the signal  $y_3, y_4$ , and the estimator predicts again the correct phase difference (bottom row).

## V. INFLUENCE OF DIFFERENT PARAMETERS ON THE IDENTIFICATION

In this section, we explore the effect of the input sensor location on the performance of the estimator. This is significantly influenced by other parameters: (i) the number of POD modes ( $k$ ), (ii) the order of the reduced-order system ( $N_x$ ), and (iii) the particular velocity component(s) recorded by the sensor ( $u, v$ , or both). We systematically study the effect of these parameters on the performance of the model. The results are grouped for models composed of  $k = 2$  (Sec.V A) and  $k = 4$  POD modes (Sec. V B).

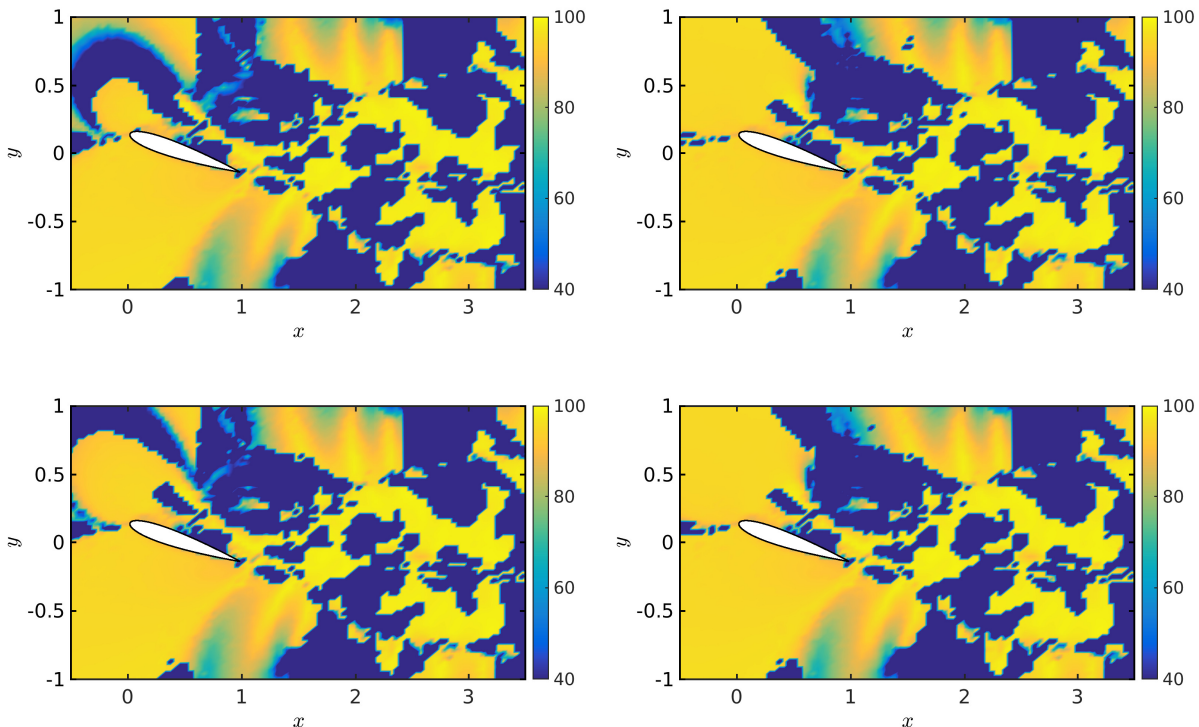


FIG. 15: FIT[%] between the actual first pair of POD coefficients and the coefficients estimated by models composed of 2 POD modes and 2 states obtained from different learning datasets. The datasets lengths are (top) 1000 and (bottom) 2000 snapshots, respectively. Left and right columns correspond to different time periods. The input signal is  $s = u$ .

### A. 2 POD modes

We consider first models with input the streamwise velocity recorded at different positions, i.e.  $s = u(x_s, y_s)$ . The number of states is set to  $N_x = 2$ . Fig. 15 displays the average of the fit of the first two POD coefficients for a validation dataset of length  $m_{TD} = 750$  (initialised by  $\mathbf{Y} = \mathbf{C}^\dagger \mathbf{X}$ ). This same dataset is used throughout this section to validate different estimators. Yellow (low error) and blue (high error) regions correspond to working and non-working locations for the sensor, respectively. The top row shows the results for a dataset of length 1000 snapshots and the bottom row results for 2000 snapshots. The right column illustrates the effect on the performance when the dataset is shifted in time. Note that the lower limit of the colour bar has been set to 40% for visualisation purposes but the actual minimal values can be lower.

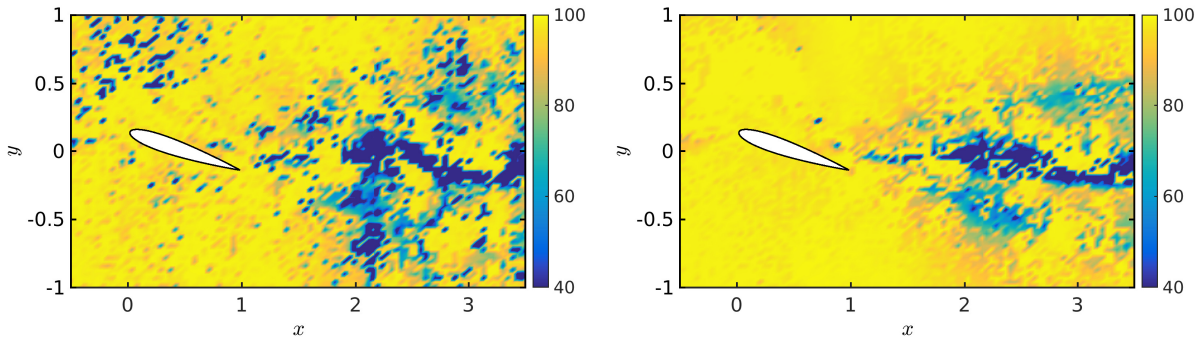


FIG. 16: Influence of the number of states of the estimator on the quality of the identification. FIT[%] between the actual first pair of POD coefficients and the coefficients estimated by models composed of 2 POD modes and (left) 4 and (right) 6 states. The input signal is  $s = u$ .

This plot shows that different datasets lead to similar distributions of working/non-working regions. The pattern downstream of the aerofoil is repeatable in all 4 datasets, except from a region upstream and above the aerofoil which is more erratic. Apart from this region, neither the length of the dataset, nor the shift in time has an impact on the performance of the model: the results are therefore robust to such changes. In the following, we fix the dataset length to 2000 snapshots.

We study next the effect of the number of states of the estimator,  $N_x$ . Fig. 16 shows the results for  $N_x = 4$  (left plot) and for  $N_x = 6$  (right plot). In comparison to  $N_x = 2$ , the utilisation of additional states leads to substantial improvement in performance, and a significant enlargement of the region where the estimator is effective. The ability to alter  $N_x$  (and thereby improve performance) is an important advantage that dynamic estimators have over static ones, which are very common in literature as explained in the Introduction.

It is interesting to notice an emerging pattern: as  $N_x$  increases, the largest error (lowest fit) is concentrated in areas that become more and more well defined along a narrow region in the wake of the aerofoil and in two patches below and above (centred around  $x \approx 2$  and 3, respectively). A physical explanation to this pattern will be provided later in this subsection.

We now explore the performance when  $s = v$  and  $s = [u, v]^T$ . The results are depicted in the left and right columns of Fig. 17, respectively, for three different values of  $N_x = 2, 4, 6$ .

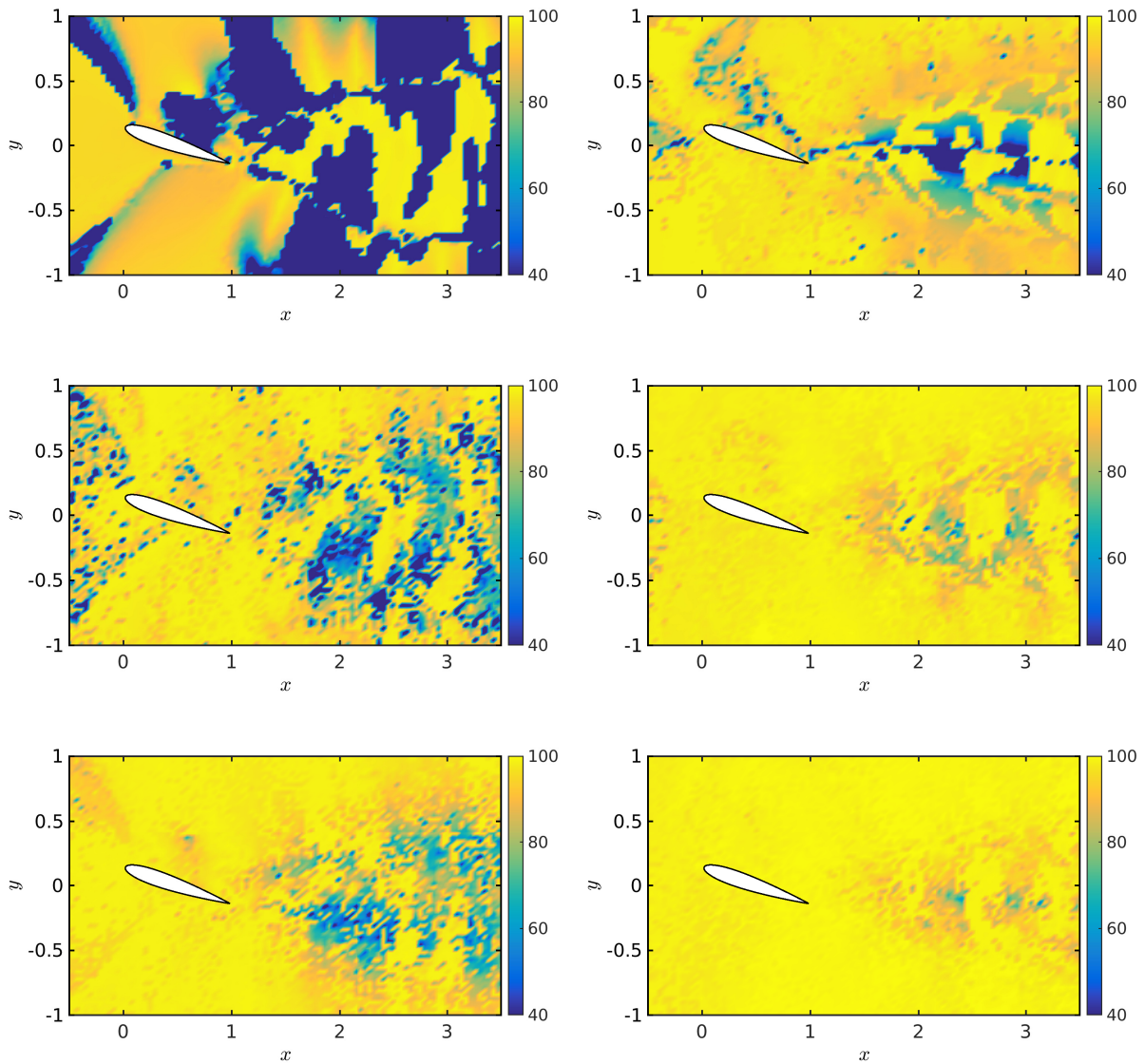


FIG. 17: Influence of the input signal on the quality of the model. FIT[%] between the actual first pair of POD coefficients and the coefficients estimated by models obtained using (left) the cross-stream  $s = v$ , and (right) both the streamwise and cross-stream  $s = [u, v]^T$  components of the oscillatory velocity. The models are composed of 2 POD modes and (top) 2, (middle) 4 and (bottom) 6 states.

Using the  $v$  velocity component appears to yield slightly improved results for all number of states investigated. As before, an increase in the number of states significantly improves the performance of the model. Again a pattern emerges, and the areas of low performance become sharper. When both velocity components are used, i.e.  $s = [u, v]^T$ , the performance

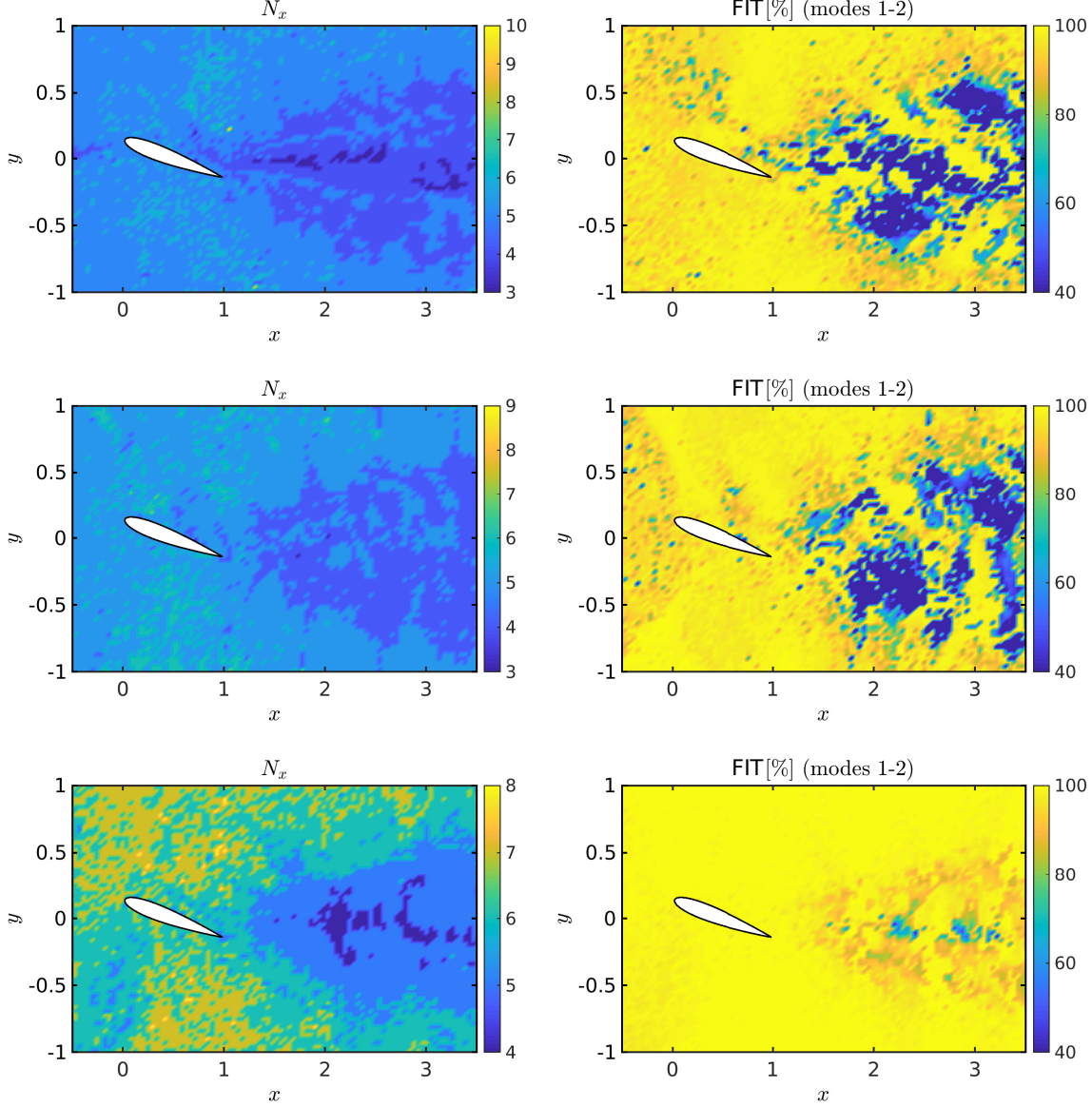


FIG. 18: (Left) Optimal number of states estimated by the N4SID algorithm for 2 POD modes and (right) FIT[%] between the actual first pair of POD coefficients and the coefficient estimated by such models. The models are obtained using (top) the streamwise  $s = u$ , (middle) the cross-stream  $s = v$  and (bottom) both the streamwise and cross-stream  $s = [u, v]^T$  components of the oscillatory velocity.

is much improved for all orders  $N_x$ . Now almost any point can be used for flow reconstruction.

It is clear that an increase of the number of states of the system leads to an increase of the estimator's performance. However, there exists a limit beyond which any additional state leads to unnecessary complex models (over-fitting). Different criteria can be used

to define if a model is under-fitted or over-fitted (see chapter 16 of the book Ljung [25]). Here, we use the optimal number of states of the model. As mentioned in Sec. II C, the N4SID algorithm can return not only the matrices that best represent the input-output behaviour of the training dataset, but also the optimal order of the estimator (note that the MATLAB implementation limits this value to 10). The order  $N_x$  is obtained from the distribution of the singular values of the weighted matrix (constructed by the algorithm using the Hankel matrices of the input-output data). The selection of the dominant singular values is straightforward when the largest ones are well separated from the rest, but it is not so clear-cut when the decay is smooth.

Fig. 18 illustrates the spatial distribution of the optimal  $N_x$  value (in the left column) and the corresponding FIT[%] (in the right column). Each row represents different input ( $s = u$ ,  $s = v$  and  $s = [u, v]^T$  for the top, middle and bottom rows respectively). When  $s = u$ , the optimal number of states is between 3 – 6, with the smaller numbers localised in the wake. In agreement with the previous figure 16, the low fit areas in the wake are again localised along a horizontal region centred around  $y \approx 0$  and two patches above and below. This result suggests that an increase in the number of states does not yield a significant change in the low fit areas. It can however affect the FIT[%] values. Indeed the fit is better in Fig. 16 compared to Fig. 18 because  $N_x$  was higher. This indicates that the improvement is probably due to over-fitting. When  $s = v$ , again the optimal  $N_x$  values are smaller in the wake, and the fit pattern is similar to that of the previous figure 17 where  $N_x$  was constant.

When  $s = [u, v]^T$ , the results indicate that a model with 4 to 5 states in the wake performs very well. There are some localised spots with low fit, centred around  $y \approx 0$ , but at these points the computed number of states is minimum (equal to  $N_x = 4$ ). It is likely that the algorithm cannot compute accurately the optimal value of  $N_x$  at these points, because there is no clear separation in the distribution of singular values (and the boundary between optimal fitting and over-fitting is not clear).

To gain further insight into these results, the spatial distribution of the time-average Reynolds stresses  $\overline{u^2}$  and  $\overline{v^2}$  of the mode pairs 1-2 and 3-4 are computed. The left column of Fig. 19 depicts the results for the projection on modes 1-2 (shedding mode) and the right column on modes 3-4 (first harmonic). This figure collects together the areas of spatial dominance of the two most significant mode pairs.

Since each pair of modes is associated with one frequency, Fig. 19 can be used to deduce

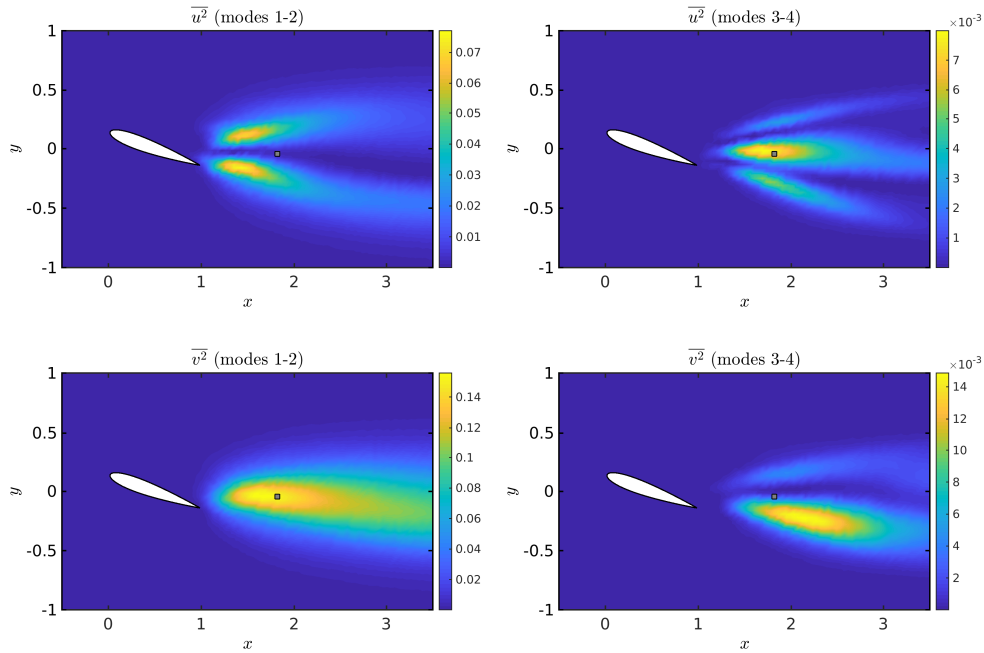


FIG. 19: Reynolds stress of the (top) streamwise and (bottom) cross-stream oscillatory velocity projected on (left) the first and second and (right) the third and fourth POD modes, respectively. Point  $(x_s, y_s) = (1.82, -0.04)$  is denoted with an open square.

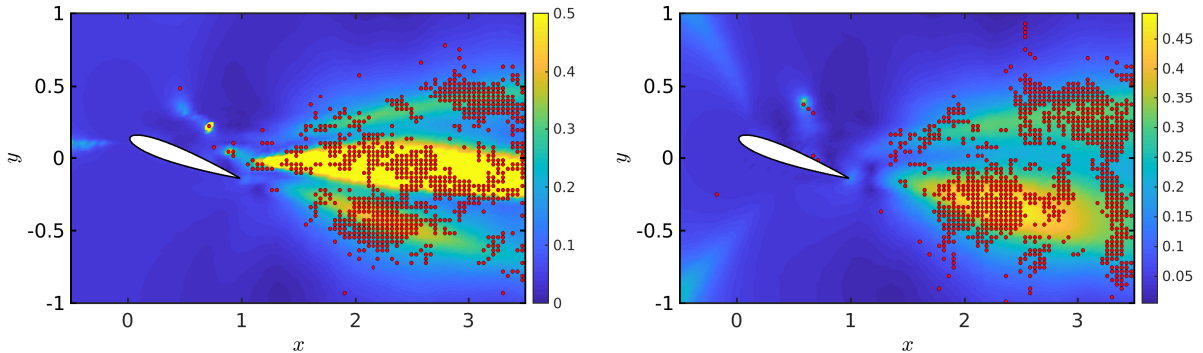


FIG. 20: Square root of the ratio of Reynolds stresses contained in modes  $3 - m$  and  $1 - 2$  superimposed with the locations of the sensor where the FIT[%] is less than 60% (shown as red dots). (left)  $\overline{u^2}$  Reynolds stress, (right)  $\overline{v^2}$  Reynolds stress.

the dominant frequencies of the signal  $s$ . For example, if the input is the streamwise  $u$  velocity (i.e.  $s = u$ ), the vortex shedding frequency ( $f = 0.6$ ) is dominant for sensors located along the path of the vortices shed from the pressure and suction sides (top, left figure). If,

on the other hand, the  $u$  sensor is located in the area sandwiched between these two paths, this main frequency is much weaker, while the first harmonic ( $f = 1.2$ ) becomes very strong. The spatial distribution of  $\overline{v^2}$  is different. For the main shedding mode, the maximum is found around  $y \approx 0$ , and is larger compared to  $\overline{u^2}$  (by a factor of about 2). If  $s = v$ , then a sensor located in this region will give a very strong signal at the main frequency. For the first harmonic frequency, the region of maximum values is slanted compared to the free-stream velocity.

For  $k = 2$ , the estimator needs to filter out any frequency different from the main shedding frequency. Therefore, we can assume that the observer’s performance will be higher when the sensor is placed in points where this main frequency is dominant. In order to verify this, we can compare the Reynolds stresses of the velocity field projected on the first pair of modes with the Reynolds stresses for the rest of them. In Fig. 20 (left), we show the ratio of Reynolds stresses for modes 3 to  $m$  and modes 1-2  $\left[\overline{u^2}\right]_{3-m} / \left[\overline{u^2}\right]_{1-2}$  and we superimpose the locations where FIT[%] is less than 60 % (as red dots). Although there is some scatter, the correlation is clear: in the three regions where the aforementioned ratio is high, the fit is low. This makes intuitive sense: it’s difficult for the model to extract the main frequency when the content at this frequency is low and the input signal is dominated by other frequencies. Although in Fig. 20, the ratio is limited to 0.5, the actual values are higher.

## B. 4 POD modes

We now turn our focus to models with  $k = 4$  POD modes. The task of estimation now becomes more difficult. The algorithm is provided with the time signal at just one point in the wake, and must extract the vector of 4 POD coefficients  $\mathbf{Y}_e = [y_{1,e}, y_{2,e}, y_{3,e}, y_{4,e}]^T$ . As before, we assessed the effects of dataset length and time shift to make sure that the results are independent of the choice of the learning dataset. Indeed, we confirmed that this is the case (results not shown for brevity). Additionally, we confirmed that, as for models with  $k = 2$  POD modes, adding more states leads to better performance of the estimators. For conciseness, the influence of the sensor location is directly illustrated using models composed of the optimal number of states only.

Fig. 21 shows the optimum number of states (left column) as well as the corresponding fits for the first and third coefficients (middle and right columns, respectively). For a model

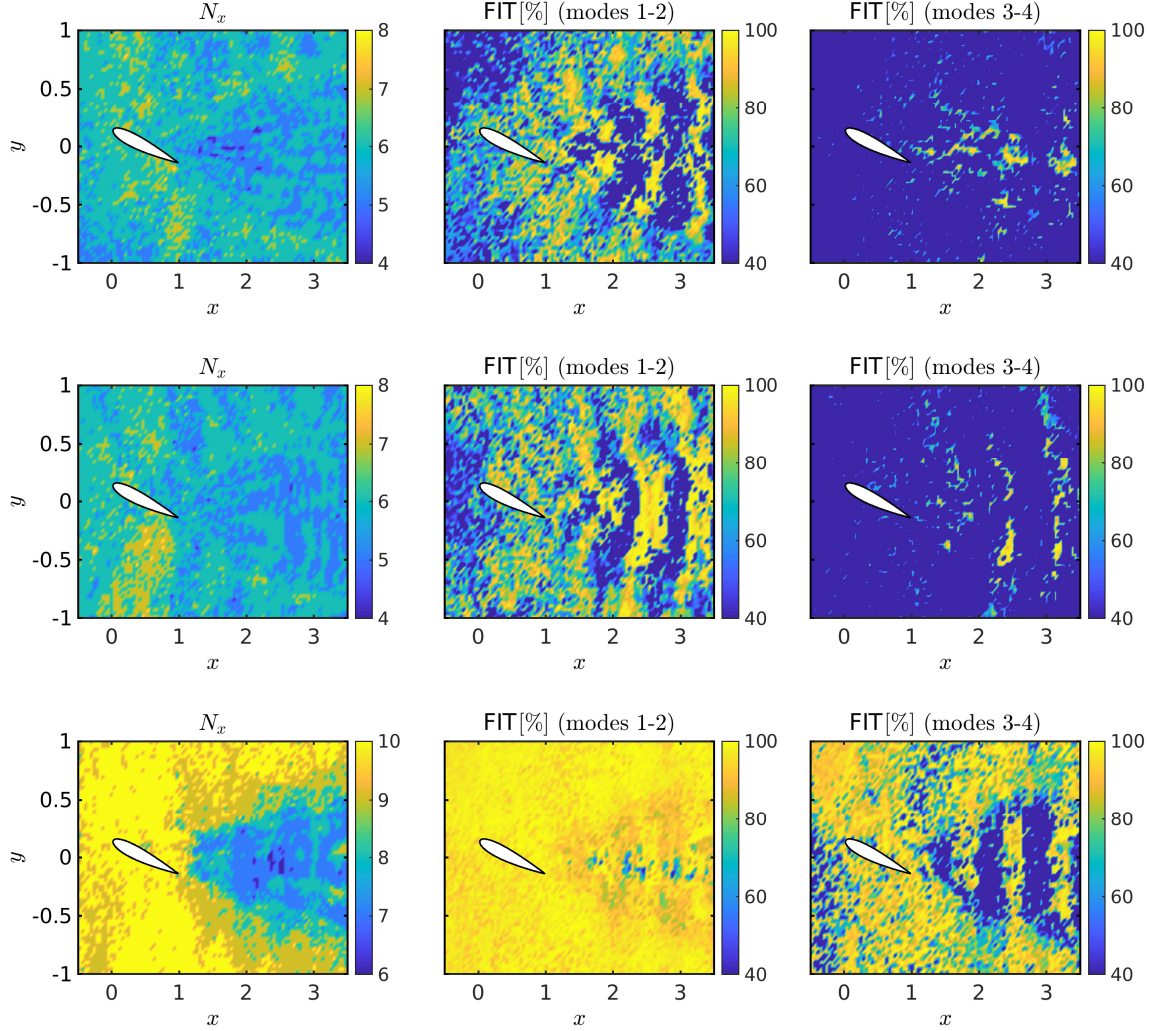


FIG. 21: (Left) Optimal number of states estimated by the N4SID algorithm for 4 POD modes and (right) FIT[%] for such models. The models are obtained using (top) the streamwise  $s = u$ , (middle) the cross-streamwise  $s = v$  and (bottom) both the streamwise and cross-streamwise  $s = [u, v]$  components of the oscillatory velocity, respectively.

utilising solely the  $u$  or  $v$  velocity components, the algorithm selects a lower number of states in the aerofoil wake resulting in relatively low fit; a similar result was also found for a model with 2 POD modes (see Fig. 18). A higher number of states is selected on top and bottom of the aerofoil. The fit is spotty, with no clear pattern. Overall, few sensor locations achieve a successful estimation when only the streamwise or cross-stream component of the velocity are considered. The fit is slightly better when the model uses the  $v$  velocity component.

When both velocity components are used, the lower number of states clearly demarcates the wake. The fit for the first mode has high values in almost every point in the domain;

there are only a few localised spots in the wake with low values. This pattern is very similar to that of Fig 18 (bottom). The coefficient of the third mode is much better captured, but a stripy fit pattern is detected.

Van Overschee and De Moor [34] realised that the N4SID algorithm is sensitive to the scaling of inputs and outputs. To explore this effect, we normalised the inputs-outputs of the system so that their variance was unity. The results obtained for the normalised case are very similar to the ones reported in Fig. 21, showing that in this case the algorithm is insensitive to scaling (results not shown for brevity).

In a further attempt to improve the performance of the estimator, we have also computed models which output the third and fourth POD coefficients only, i.e.  $\mathbf{Y}_e = [y_{e,3}, y_{e,4}]$ . We know that any interaction between the main shedding mode and the harmonics is non-linear. The estimator, being linear, can only capture this interaction through the sensor. The outputs 1-2 and 3-4 of the estimator can therefore be decoupled and can be represented by independent estimators.

Fig. 22 shows the optimal number of states (left column) and the corresponding fit for the third coefficient (right column). A clear improvement is obtained for models using solely  $u$  and  $v$ . The working areas are now larger and better defined compared to Fig. 21. Additionally, we observe that the working/non-working locations present a similar pattern to the ones for modes 1-2 (Fig. 18), but inverted. This can be explained again by reference to Fig. 20; when the signature of the main frequency in the sensor is very strong compared to the one of the first harmonic, it is more difficult for the estimator to filter it out. For estimators using both components of the velocity field, a stripy pattern is detected again. Although there is a small improvement, the fit distribution shows a very similar pattern to the one obtained for the the coupled case (Fig. 21).

## VI. PERFORMANCE ANALYSIS AT OFF-DESIGN CONDITIONS

In this section, we explore the behaviour of estimators at off-design conditions. To this end, we performed two additional simulations at  $Re = 550$  and  $650$  ( $\pm 10\%$  variation with respect to the baseline case). At these Reynolds numbers, the characteristic frequency of the main mode shifts from  $f = 0.61$  (corresponding to  $Re = 500$ ) to  $f = 0.6$  and  $0.625$ , respectively. Reduced-order datasets are obtained from these simulations by subtracting the

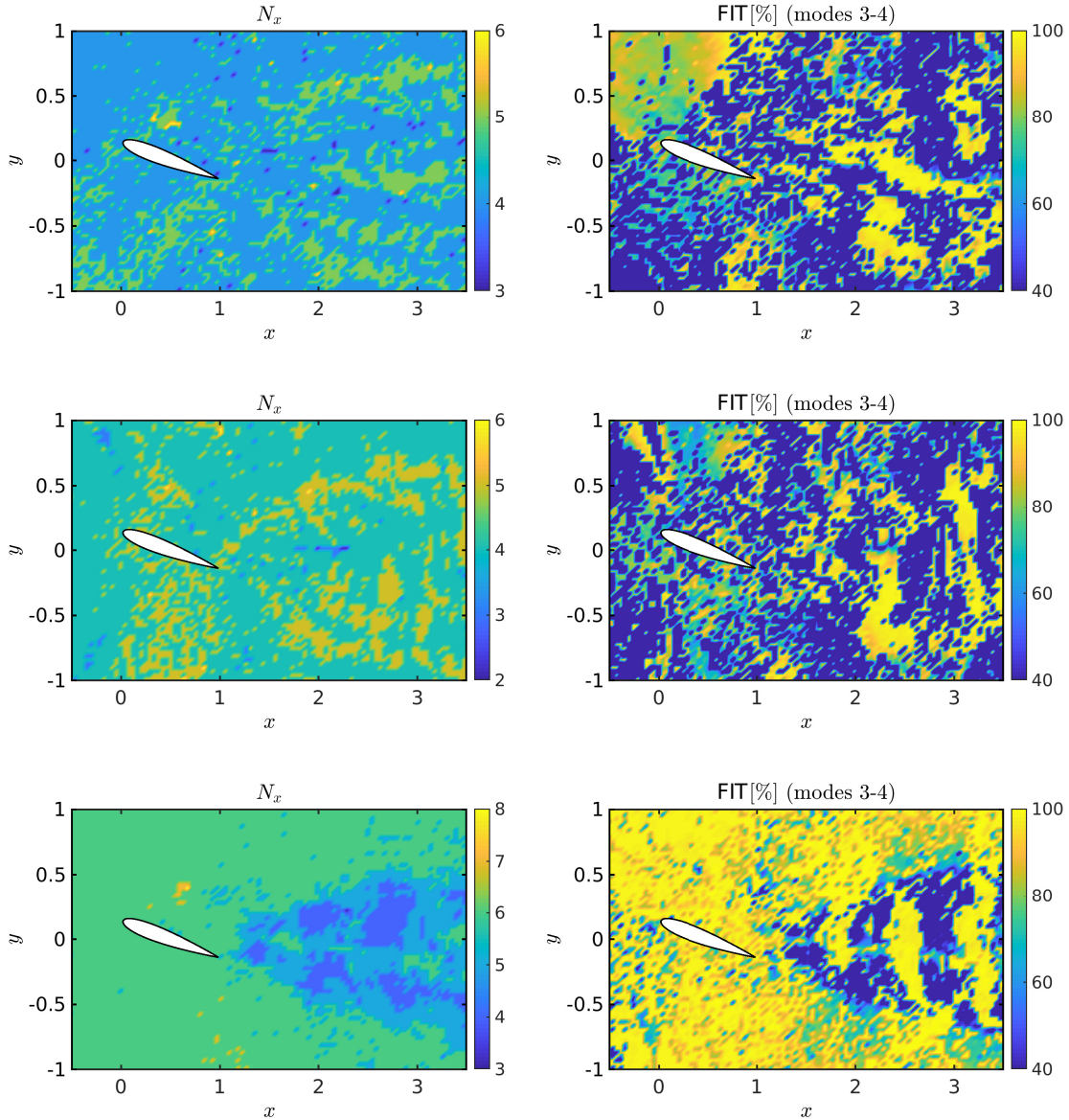


FIG. 22: (Left) Optimal number of states estimated by the N4SID algorithm for the 3rd and 4th POD modes and (right) FIT[%] for such models. The models are obtained using (top) the streamwise  $s = u$ , (middle) the cross-streamwise  $s = v$  and (bottom) both the streamwise and cross-streamwise  $s = [u, v]^T$  components of the oscillatory velocity, respectively.

mean flow and projecting the perturbation field onto the POD modes. Both the mean flow and the POD modes used here are obtained at  $Re = 500$ . A time-segment of each dataset of length  $m_{TD} = 750$  is then supplied into the models extracted in the previous sections to assess their robustness.

We focus our analysis on the set of parameters that provided the best results in section V, i.e. input sensors composed of both the streamwise and cross-stream components of the velocity, decoupled models for the first (modes 1-2) and second (modes 3-4) pairs of POD modes and optimal number of states. These conditions correspond to the estimators depicted in the bottom rows of Figs. 18 and 22, respectively. Fig. 23 shows the goodness of fit between the true off-design testing datasets and the predictions of models obtained with sensors located at different positions. We observe an overall loss of performance with respect to the nominal case, with similar trends for  $Re = 550$  and  $650$ . Estimation of the first pair of POD coefficients is successful for sensors placed in many locations in the wake of the aerofoil, but it fails in the regions situated above, below and upstream of it. This is in contrast to the baseline case (Fig. 18), where the estimators work for most of the sensor locations. For the second pair of coefficients, we observe a similar trend. A common feature of these non-working areas is the low amplitude of the input signal, which requires the estimators to have very large gains  $\mathcal{O}(10^{4-5})$  around the frequency of interest. These estimators will strongly amplify any uncertainty of the input signal in that frequency range leading to the diminished robustness observed in Fig. 23.

The influence of the number of states on the quality of the estimation for off-design conditions was also studied. The results obtained for  $N_x = 2, 4$  and  $6$  show similar trends to those observed in Fig. 23 and are not shown for brevity.

We now assess the robustness of the estimator in the full-order space (velocity field). To this end, the location of the input sensor is fixed to  $(x_s, y_s) = (2.45, -0.10)$ , both components of the velocity are used as input and two independent, decoupled models are obtained for the first and second pairs of POD coefficients. For these conditions, the vector of fits for the learning dataset ( $Re = 500$ ) is  $\text{FIT}[\%] = [99.19, 98.69, 97.09, 97.62]^\top$ . For a validation dataset measured at  $Re = 550$  the performance is slightly reduced, that is  $\text{FIT}[\%] = [82.68, 81.87, 83.11, 83.75]^\top$ . The perturbation velocity field is then reconstructed using Eq. 8(a) and the POD modes obtained at  $Re = 500$ . Fig. 24 shows the original and reconstructed velocity at two arbitrary time instants. Again, the vortex shedding is clearly captured by the model and the reconstructions are very similar. Some discrepancies can be observed that are linked to the reduced-basis: the first four POD modes computed at  $Re = 500$  span a slightly different subspace than the one span by the main and first harmonic modes at any other Reynolds number.

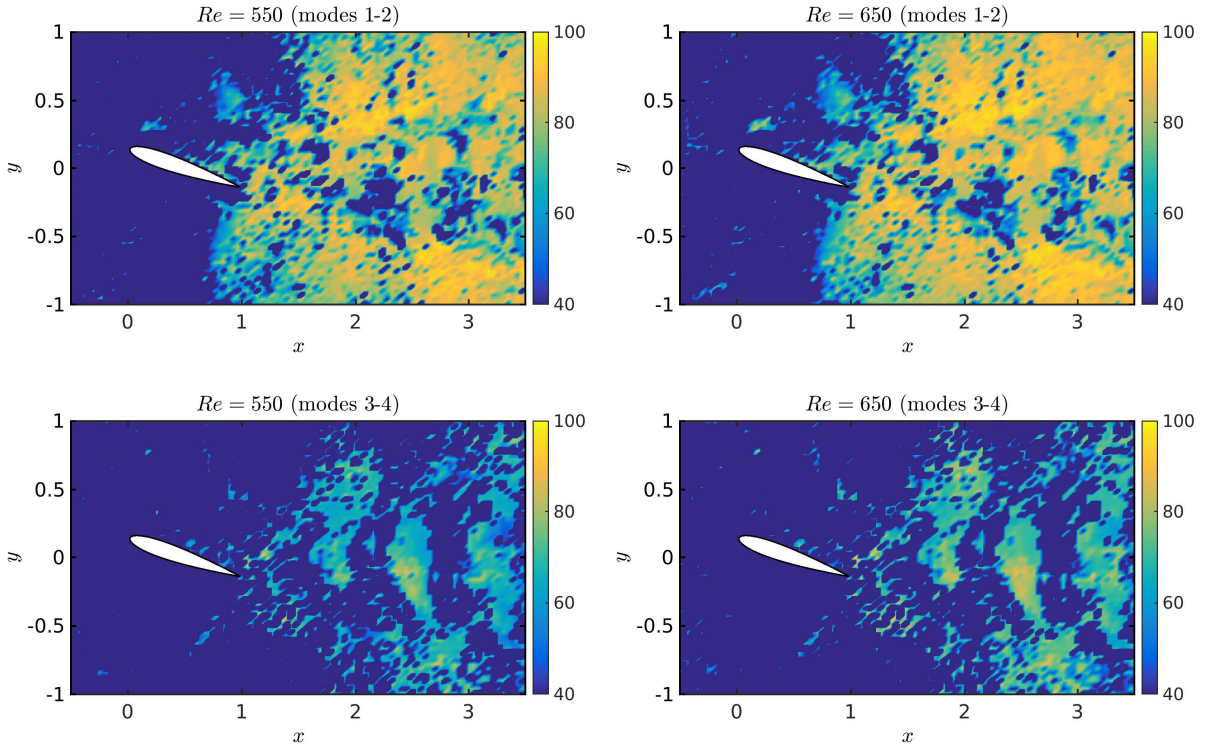


FIG. 23: Performance of the estimator at off-design conditions: models extracted from data obtained at  $Re = 500$  are evaluated with datasets collected at (left)  $Re = 550$  and (right)  $Re = 650$ . FIT[%] between the actual (top) first and (bottom) second pair of POD coefficients and the predictions of the model. The models explicitly decouple the first and second pair of POD modes and use the optimal number of states given in the bottom row of Figs. 18 and 22 for the first and second pair of coefficients, respectively. We have used

$$s = [u, v]^T.$$

## VII. SUMMARY AND CONCLUSIONS

The central objective of this paper was to derive a data-driven, dynamic observer that can reconstruct the full 2D velocity field from a single point measurement in the wake of an aerofoil. In the present work, the data were produced by running a CFD simulation, but equally well they could have been obtained using PIV. To reduce the spatial dimensionality of the data, we constructed a reduced order model based on POD.

The system identification algorithm, N4SID, was then applied in conjunction with the dynamic observer to produce the final model. The effect of a number of parameters on the performance of the algorithm was assessed and the results were related to the flow dynamics

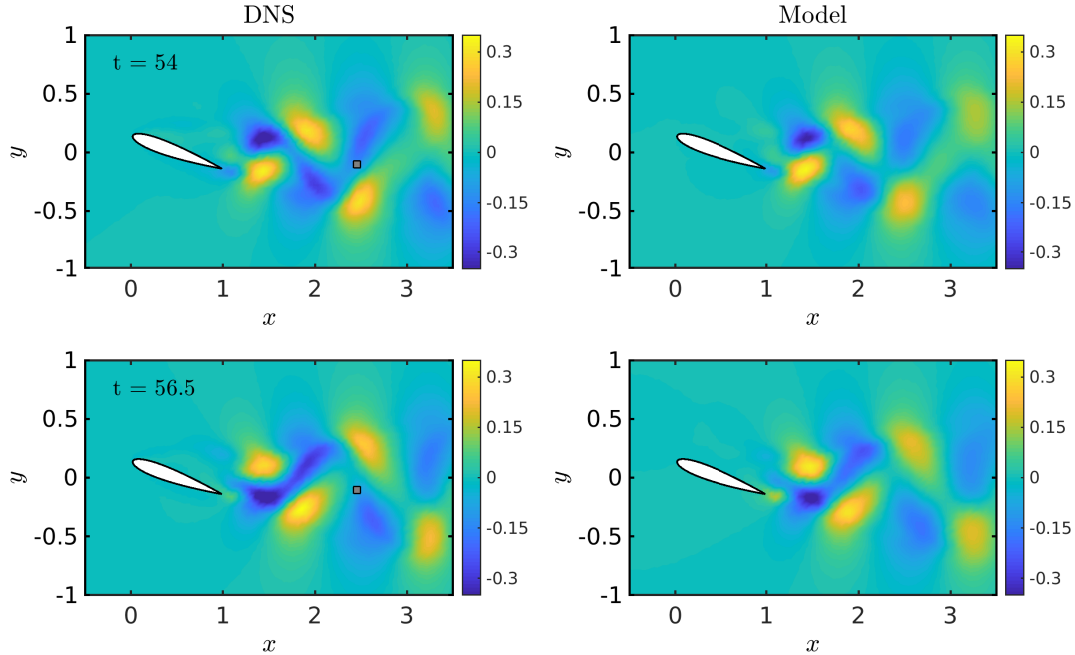


FIG. 24: Streamwise component of the (left) true oscillatory velocity obtained from a simulation at  $Re = 550$  and (right) the prediction of a model extracted from data measured at  $Re = 500$ . The grey squares represent the position of the estimation sensor.

and mode interaction in the wake. More specifically, the number of POD modes, length of learning dataset, input velocity component(s), observer initialisation, sensor location and number of states of the estimator, were all investigated. Even when initialising the testing dataset with a state of  $\mathbf{X}_e(t) = 0$ , the algorithm was able to estimate the underlying coefficients of the model, though with an added delay. Average fit values of over 90% were shown to be achievable, especially for the first and second POD modes, that capture most of the energy. Best results are obtained when both  $u$  and  $v$  velocity components are measured.

Overall the algorithm was found to achieve the reconstruction of flow fields with an impressive degree of accuracy. Considering the unsteady and non-linear nature of the flow, this is especially surprising and opens the approach up to a wide variety of applications [44].

As a result of the analysis described in this paper we can provide some general guidelines for application of the method. First, we recommend the use of the single-point estimator for the reconstruction only of the main shedding mode. In this case, estimators are found to work using measurements from most of the wake locations for both design and off-design conditions. For the reconstruction of the harmonics, the use of independent estimators for

each harmonic is the best option in a linear-framework. If the harmonics can be assumed slaves of the main mode [22], they might be modelled using non-linear triadic interaction [22] together with the linear estimator for the main mode (this approach is outside of the scope of the current publication). Best estimation results will be obtained using both components of the velocity. To obtain estimators that are robust to model uncertainties, the estimation sensor should be placed in regions where the amplitude of the main mode is high to avoid large gains in the transfer functions.

- 
- [1] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, 2nd ed., Cambridge Monographs on Mechanics (Cambridge University Press, 2012).
  - [2] K. Willcox, “Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition,” *Computers & Fluids* **35**, 208 – 226 (2006).
  - [3] R. Everson and L. Sirovich, “Karhunen–loève procedure for gappy data,” *J. Opt. Soc. Am. A* **12**, 1657–1664 (1995).
  - [4] B. Yildirim, C. Chrysostomidis, and G.E. Karniadakis, “Efficient sensor placement for ocean measurements using low-dimensional concepts,” *Ocean Modelling* **27**, 160 – 173 (2009).
  - [5] R. Semaan, “Optimal sensor placement using machine learning,” *Computers & Fluids* **159**, 167 – 176 (2017).
  - [6] B. J. McKeon and A. S. Sharma, “A critical-layer framework for turbulent pipe flow,” *J. Fluid Mech* **658**, 336–382 (2010).
  - [7] S. Beneddine, R. Yegavian, D. Sipp, and B. Leclaire, “Unsteady flow dynamics reconstruction from mean flow and point sensors: an experimental study,” *J. Fluid Mech* **824**, 174–201 (2017).
  - [8] F. Gómez, A. S. Sharma, and H. M. Blackburn, “Estimation of unsteady aerodynamic forces using pointwise velocity data,” *J. Fluid Mech* **804** (2016).
  - [9] N. Thomareis and G. Papadakis, “Resolvent analysis of separated and attached flows around an airfoil at transitional reynolds number,” *Phys. Rev. Fluids* **3**, 073901 (2018).
  - [10] S. J. Illingworth, J. P. Monty, and I. Marusic, “Estimating large-scale structures in wall turbulence using linear models,” *J. Fluid Mech* **842**, 146–162 (2018).
  - [11] S. Brunton, J. Tu, I. Bright, and J. Kutz, “Compressive sensing and low-rank libraries

- for classification of bifurcation regimes in nonlinear dynamical systems,” *SIAM Journal on Applied Dynamical Systems* **13**, 1716–1732 (2014).
- [12] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, “Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns,” *IEEE Control Systems Magazine* **38**, 63–86 (2018).
- [13] R. J. Adrian, “Conditional eddies in isotropic turbulence,” *Physics of Fluids* **22**, 2065–2070 (1979).
- [14] R. J. Adrian and P. Moin, “Stochastic estimation of organized turbulent structure: homogeneous shear flow,” *J. Fluid Mech* **190** (1988).
- [15] Y.G. Guezennec, “Stochastic estimation of coherent structures in turbulent boundary layers,” *Phys. Fluids* **1**, 1054–1060 (1989).
- [16] J.-P. Bonnet, D.R. Cole, J. Delville, M.N. Glauser, and L.S. Ukeiley, “Stochastic estimation and proper orthogonal decomposition: complementary techniques for identifying structure,” *Exp. Fluids* **17**, 307–314 (1994).
- [17] J. A. Taylor and M. N. Glauser, “Towards practical flow sensing and control via pod and lse based low-dimensional tools,” *Journal of fluids engineering* **126**, 337–345 (2004).
- [18] C.E. Tinney, F. Coiffet, J. Delville, A.M. Hall, P. Jordan, and M.N. Glauser, “On spectral linear stochastic estimation,” *Exp. Fluids* **41**, 763–775 (2006).
- [19] J. H Tu, J. Griffin, A. Hart, C. W. Rowley, L. N. Cattafesta, and L. S. Ukeiley, “Integration of non-time-resolved piv and time-resolved velocity point sensors for dynamic estimation of velocity fields,” *Experiments in fluids* **54**, 1429 (2013).
- [20] H. Clark, A. Naghib-Lahouti, and P. Lavoie, “General perspectives on model construction and evaluation for stochastic estimation, with application to a blunt trailing edge wake,” *Experiments in fluids* **55**, 1756 (2014).
- [21] T. Kailath, B. Hassibi, and A. H. Sayed, *Linear estimation* (Prentice-Hall International, 2000).
- [22] J. Gong, J. P. Monty, and S. J. Illingworth, “Model-based estimation of vortex shedding in unsteady cylinder wakes,” *arXiv preprint arXiv:1905.00133* (2019).
- [23] D. F. Gomez, F. D. Lagor, P. B. Kirk, A. H. Lind, A. R. Jones, and D. A. Paley, “Data-driven estimation of the unsteady flowfield near an actuated airfoil,” *Journal of Guidance, Control, and Dynamics* , 1–9 (2019).

- [24] R. F. Stengel, *Optimal Control and Estimation* (Dover Publications Inc, 2nd edition, New York, 1994).
- [25] L. Ljung, *System identification: theory for the user* (Prentice-Hall PTR, 1999).
- [26] A. Hervé, D. Sipp, P. J. Schmid, and M. Samuelides, “A physics-based approach to flow control using system identification,” *J. Fluid Mech* **702**, 26–58 (2012).
- [27] F. Juillet, P. J. Schmid, and P. Huerre, “Control of amplifier flows using subspace identification techniques,” *J. Fluid Mech* **725**, 522–565 (2013).
- [28] F. Juillet, B. J. McKeon, and P. J. Schmid, “Experimental control of natural perturbations in channel flow,” *J. Fluid Mech* **752**, 296–309 (2014).
- [29] N. Gautier and J. L. Aider, “Feed-forward control of a perturbed backward-facing step flow,” *J. Fluid Mech* **759**, 181–196 (2014).
- [30] J. Guzmán-Iñigo, D. Sipp, and P. J. Schmid, “A dynamic observer to capture and control perturbation energy in noise amplifiers,” *J. Fluid Mech* **758**, 728–753 (2014).
- [31] J. Guzmán-Iñigo, D. Sipp, and P. J. Schmid, “Recovery of the inherent dynamics of noise-driven amplifier flows,” *J. Fluid Mech* **797**, 130–145 (2016).
- [32] P. Van Overschee and B. De Moor, “N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems,” *Automatica* **30**, 75–93 (1994).
- [33] S.J. Qin, “An overview of subspace identification,” *Comp. & Chem. Eng.* **30**, 1502–1513 (2006).
- [34] P. Van Overschee and B.L. De Moor, *Subspace Identification for Linear Systems; Theory — Implementation — Applications* (Springer US, 1996).
- [35] J. C. Loiseau, B. R. Noack, and S. L. Brunton, “Sparse reduced-order modelling: sensor-based dynamics to full-state estimation,” *J. Fluid Mech* **844**, 459–490 (2018).
- [36] M. Korda and I. Mezić, “Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control,” *Automatica* **93**, 149–160 (2018).
- [37] S. Chaturantabut and D. C. Sorensen, “Nonlinear model reduction via discrete empirical interpolation,” *SIAM Journal on Scientific Computing* **32**, 2737–2764 (2010).
- [38] M. Fosas de Pando, P. J. Schmid, and D. Sipp, “Nonlinear model-order reduction for compressible flow solvers using the discrete empirical interpolation method,” *J. Comp. Phys.* **324**, 194–209 (2016).
- [39] R. Penrose, “A generalized inverse for matrices,” *Mathematical proceedings of the Cambridge philosophical society* **51**, 406–413 (1955).

- [40] A.C. Antoulas, *Approximation of large-scale dynamical systems* (SIAM Publishing, 2005).
- [41] P. Van Overschee and B. De Moor, “A unifying theorem for three subspace system identification algorithms,” *Automatica* **31**, 1853–1864 (1995).
- [42] W. Zhang and R. Samtaney, “Biglobal linear stability analysis on low-re flow past an airfoil at high angle of attack,” *Physics of Fluids* **28**, 044105 (2016).
- [43] D. Sipp and A. Lebedev, “Global stability of base and mean flows: a general approach and its applications to cylinder and open cavity flows,” *J. Fluid Mech* **593**, 333–358 (2007).
- [44] C. Leclercq, F. Demourant, C. Poussot-Vassal, and D. Sipp, “Linear iterative method for closed-loop control of quasiperiodic flows,” *J. Fluid Mech* **868**, 26–65 (2019).